3D reconstruction for plastic surgery simulation based on statistical shape models

Guillermo Ruiz Fernández

DOCTORAL THESIS UPF / YEAR 2018

THESIS SUPERVISORS
Dr. Miguel Ángel González Ballester [1,2]
Dr. Federico Mateo Sukno [1]

[1] Department of Information and Communication Technologies
[2] ICREA, Barcelona, Spain

Hoy es siempre todavía,
toda la vida es ahora.

Antonio Machado

A mi padre, Jose Manuel y a mi madre, Esther. Gracias.

Acknowledgments

The decision of doing a Ph.D. supposes a great personal challenge. Along the way you go through all kind of emotions. Sometimes you feel satisfaction with the obtained results and the work done. Other times, when something does not go well you feel frustration. It is not always easy. There are a number of people who in one way or another have contributed to make this adventure more enjoyable by giving their support and help. I wanted to thank all of them.

First of all, I would like to thank my supervisors Miguel Ángel González and Federico Sukno for giving me the opportunity of being part of the department of Information and Communication Technologies in the UPF. Thank you for all of your advices, your guidance and everything you taught me during these years.

I would like to thank Jaime García, for giving me the great opportunity of joining Crisalix, providing the possibility to carry out this Ph.D. in the company. It has been a really nice experience. I am very grateful to have been part of the development team during these years where I have learned many things. I would like to especially thank Edu, for being a great colleague in the master that we studied together, and even a better work colleague afterwards. Thank you for all your always good advices and all your invaluable help. It is a pleasure to work by your side. Pol, many thanks for your willingness to help and always have the best advices. Thank you for your awesome tools and your insistence to make me write using pure Latex. Many thanks to the rest of my Crisalix colleagues, for being part of this great work atmosphere and because from all of you I keep learning new things day by day.

Finally, I would like to give my most deepest and sincere gratitude to the people that most matter to me. To my father, my mother and my sister for all their love. Thanks for giving me the opportunities and experiences that have made me who I am. Thank you for always being there. And of course thank you Adri, for being by my side every day, bringing me your unconditional support and love.

Abstract

This thesis has been accomplished in Crisalix in collaboration with the Universitat Pompeu Fabra within the program of Doctorats Industrials. Crisalix has the mission of enhancing the communication between professionals of plastic surgery and patients by providing a solution to the most common question during the surgery planning process of "How will I look after the surgery?". The solution proposed by Crisalix is based in 3D imaging technology. This technology generates the 3D reconstruction that accurately represents the area of the patient that is going to be operated. This is followed by the possibility of creating multiple simulations of the plastic procedure, which results in the representation of the possible outcomes of the surgery.

This thesis presents a framework capable to reconstruct 3D shapes of faces and breasts of plastic surgery patients from 2D images and 3D scans. The 3D reconstruction of an object is a challenging problem with many inherent ambiguities. Statistical model based methods are a powerful approach to overcome some of these ambiguities. We follow the intuition of maximizing the use of available prior information by introducing it into statistical model based methods to enhance their properties.

First, we explore Active Shape Models (ASM) which are a well-known method to perform 2D shapes alignment. However, it is challenging to maintain prior information (*e.g.* small set of given landmarks) unchanged once the statistical model constraints are applied. We propose a new weighted regularized projection into the parameter space which allows us to obtain shapes that, at the same time, fulfill the imposed shape constraints and are plausible according to the statistical model.

Second, we extend this methodology to be applied to 3D Morphable Models (3DMM), which are a widespread method to perform 3D reconstruction. However, existing methods present some limitations. Some of them are based in non-linear optimizations computationally expensive that can get stuck in local minima. Another limitation is that not all the methods provide enough resolution to represent accurately the anatomy details needed for this application. Given the medical use of the applica-

tion, the accuracy and robustness of the method, are important factors to take into consideration. We show how 3DMM initialization and 3DMM fitting can be improved using our weighted regularized projection.

Finally, we present a framework capable to reconstruct 3D shapes of plastic surgery patients from two possible inputs: 2D images and 3D scans. Our method is used in different stages of the 3D reconstruction pipeline: shape alignment; 3DMM initialization and 3DMM fitting. The developed methods have been integrated in the production environment of Crisalix, proving their validity.

Resum

Aquesta tesi ha estat realitzada a Crisalix amb la col·laboració de la Universitat Pompeu Fabra sota el pla de Doctorats Industrials. Crisalix té com a objectiu la millora de la comunicació entre els professionals de la cirurgia plàstica i els pacients, proporcionant una solució a la pregunta que sorgeix més freqüentment durant el procés de planificació d'una operació quirúrgica "Com em veuré després de la cirurgia?". La solució proposada per Crisalix està basada en la tecnologia d'imatge 3D. Aquesta tecnologia genera la reconstrucció 3D de la zona del pacient operada, seguit de la possibilitat de crear múltiples simulacions obtenint la representació dels possibles resultats de la cirurgia.

Aquesta tesi presenta un sistema capaç de reconstruir cares i pits de pacients de cirurgia plàstica a partir de fotos 2D i escanegis. La reconstrucció en 3D d'un objecte és un problema complicat degut a la presència d'ambigüitats. Els mètodes basats en models estadístics son adequats per mitigar-les. En aquest treball, hem seguit la intuïció de maximitzar l'ús d'informació prèvia, introduint-la al model estadístic per millorar les seves propietats.

En primer lloc, explorem els Active Shape Models (ASM) que són un conegut mètode fet servir per alinear contorns d'objectes 2D. No obstant, un cop aplicades les correccions de forma del model estadístic, es difícil de mantenir informació de la que es disposava a priori (per exemple, un petit conjunt de punts donat) inalterada. Proposem una nova projecció ponderada amb un terme de regularització, que permet obtenir formes que compleixen les restriccions de forma imposades i alhora són plausibles en concordança amb el model estadístic.

En segon lloc, ampliem la metodologia per aplicar-la als anomenats 3D Morphable Models (3DMM) que són un mètode extensivament utilitzat per fer reconstrucció 3D. No obstant, els mètodes de 3DMM existents presenten algunes limitacions. Alguns estan basats en optimitzacions no lineals, computacionalment costoses i que poden quedar atrapades en mínims locals. Una altra limitació és que no tots el mètodes proporcionen la resolució adequada per representar amb precisió els detalls de l'anatomia.

Donat l'ús mèdic de l'aplicació, la precisió i la robustesa són factors molt importants a tenir en compte. Mostrem com la inicialització i l'ajustament de 3DMM poden ser millorats fent servir la projecció ponderada amb regularització proposada.

Finalment, es presenta un sistema capaç de reconstruir models 3D de pacients de cirurgia plàstica a partir de dos possibles tipus de dades: imatges 2D i escaneigs en 3D. El nostre mètode es fa servir en diverses etapes del procés de reconstrucció: alineament de formes en imatge, la inicialització i l'ajustament de 3DMM. Els mètodes desenvolupats han estat integrats a l'entorn de producció de Crisalix provant la seva validesa.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

When a person visits a doctor to undergo plastic surgery, the communication between the doctor and the patient during the surgery planning process is of vital importance. The patient's expectations have to be met after the surgery to avoid unsatisfactory results and even the need to re-operate. Therefore, it is preferable for the patients to be highly involved during the surgery planning process. Important aspects of the surgery planning, such as the implant selection or the surgical approach election, can be supported by a realistic representation of the outcome of the surgery giving the patients an approximate idea of how they will look after the surgery. Existing methods of patient assessment are two-dimensional images based. The main limitation of these methods is the lack of depth dimension, leading to inconsistencies due to the representation of three-dimensional objects in two dimensions. More recent imaging methods such as volumetric medical images (*e.g.* MRI) or 3D scanners can provide accurate three-dimensional representations of the patient's body. However, they are based on the use of huge hardware devices that might result invasive for the patients, time-consuming and very expensive. All these limitations make the aforementioned solutions inadequate to be used in medical assessments.

This dissertation has been carried out in Crisalix in collaboration with the Universitat Pompeu Fabra (UPF). The mission of Crisalix is to enhance the communication between surgeons and patients. The proposed solution to overcome the limitations of traditional patient assessment methods is based on 3D imaging technology. This technology generates the 3D model of the region of the patient that is going to be operated. Furthermore, when these 3D models are generated, they can be used to illustrate any plastic or aesthetic procedure simulation in order to let the patients visualize how they would look after having the surgery. Through a web-based application that provides a set of tools, the doctors can create multiple simulations for a patient, trying out different implants or transforming the 3D reconstruction in different ways, so that the patient can have an idea of the possible outcomes of a procedure and choose the one that suits him/her best. Also, this tool allows the doctor to advise the patients in an interactive way, helping the patients feel like they too have a say in how they will look after the surgery.

The generation of the patient's 3D model is achieved by performing 3D reconstruction. 3D reconstruction of an object consists of determining its geometry and appearance from incomplete data. This is a challenging problem with many inherent ambiguities. For this reason, it is crucial to define the input data for the reconstruction method, because it will define the complexity and the possible solutions to the problem. In addition, it is very important to think how this data can be acquired, since it will not be done by the researchers but by the doctors. In this thesis, we have explored the use of two different kinds of input data: 2D images and 3D point clouds obtained with a structured light based scan. These two different inputs are suitable to be used in an application such as the one described above. 2D images are very easy to obtain and handle with a great variety of devices. However, performing 3D reconstruction from 2D images can be a difficult challenge if there is no extra information of the scene. On the other hand, the acquisition of 3D point clouds with a structured light scan provides a robust and really valuable depth information conditioned to the use of a scanner device. Despite the need of a device, existing systems specifications are suitable to be used for this application

since they are portable and affordable.

In order to perform 3D reconstruction from images or depth data, researchers usually explore ways to minimize the ambiguities of the problem that can help to reduce its complexity and increase the robustness and accuracy of the methods. In some cases, this is done by assuming knowledge about the scene conditions, such as illumination or the acquisition of information with calibrated cameras or devices. In other approaches, statistical models or regression models are used to represent the characteristics of the reconstructed objects and are built from sets of training data previously acquired.

Model-based methods are widely used to perform 3D reconstruction. These methods use predefined templates or a statistical model, providing prior knowledge about the geometry of the reconstructed object. Furthermore model-based methods are robust against noisy data, incomplete scans or occluded elements. In particular, 3D Morphable Models (3DMM) proposed by [Blanz and Vetter, 1999], are extensively used for 3D face reconstruction. The 3DMM consists of building a statistical model by extracting the mean shape and the variation modes within a population from a database. However, the existing 3DMM methods present some limitations. Some of them are based on non-linear optimizations which are computationally expensive and can get stuck in local minima. Another limitation that some 3DMM methods have is that they are not adequate to work in uncontrolled environments. In addition, given the medical use of the application, the accuracy and robustness of the method are important factors to take into consideration.

In this thesis, we focus on the study of statistical models and 3DMM, exploring ways to extend the methodology to maximize the use of the available prior information, with the aim to enhance its properties and overcome the aforementioned limitations. The final goal is to provide a robust framework which obtains 3D reconstructions of faces and breasts from plastic surgery patients representing as much accurately as possible the real patient's anatomy.

1.1 Motivation and objectives

When 3D reconstruction is applied for medical image applications, some additional challenges arise. In a field such as aesthetic surgery, and even more when the method has to be implemented to create a commercial application for the use of professional surgeons there are important requirements that need to be addressed carefully.

In first place, the used method should be a portable solution suitable to be used in surgeons offices, therefore it should not depend on huge or expensive hardware devices, which can provide accurate 3D reconstructions but are not appropriate for the aforementioned scenario.

Another important aspect of the method is the accuracy. The patients planning to have surgery, will be able to see themselves "reflected" on the reconstructions and simulations. Therefore, it is very important that the reconstructions represent very accurately the anatomy of the patients, and the robustness of the method is essential in order to provide good quality and suitable reconstructions. Otherwise, the patients might have doubts or directly decide not to go through with the surgery.

The method should be also fast enough to be used during a medical appointment. When the doctor wants to instruct and advise the patient using the visual support of the 3D reconstruction, it is desirable that it can be provided at that specific moment without the need of long waiting periods or scheduling a second appointment.

After analyzing the application requirements and the aforementioned limitations of the existing methods, the research in this work covers the following points:

- Performing 3D reconstruction from 2D images implies the advantage of using an input easily obtainable, but is a challenging task. The 3D reconstruction from 2D images pipeline involves many tasks that have to be resolved in order to recover the 3D geometry of the reconstructed object. It is very common that the first important task to be resolved is the 2D landmarks alignment in the input images, that later will be put in correspondence to obtain the 3D information. The process of 2D landmark alignment has to be robust since

its output will determine the result of the 3D reconstruction. Active Shape Models (ASM) are a powerful and well-known method proposed by [Cootes et al., 1995] to locate boundaries of objects in images. They are based on obtaining a statistical shape model built from a collection of 2D shapes. Our first intuition is to explore the introduction of prior information into the ASM to enhance its properties. In some applications, there might be prior information available and some landmarks can be known beforehand (*i.e.*, the user is asked to manually locate a small set of landmarks). The goal is to introduce this prior information within the statistical model to increase the accuracy and robustness of the 2D shape alignment method.

- The following step is to extend the previous methodology to work with statistical models based on 3D shapes. The goal is to integrate 3D shape prior information into a statistical model built from 3D shapes. 3DMM are a robust and widely used method to perform 3D reconstruction. However, if additional prior information is combined with the 3DMM statistical model, shape constraints can be imposed to improve the 3DMM properties.

- After working with both 2D shapes and 3D shapes and exploring the intuition of introducing prior information within statistical models, the goal is to integrate these methods in Crisalix's production 3D reconstruction framework proving their validity. This framework should be able to process different inputs (2D images and scans) from plastic surgery patients. Using the proposed method, the objective is to provide robustness and increase the quality of the patients 3D reconstructions.

1.2 Thesis outline and contributions

This thesis provides not only a theoretic framework but also a set of algorithmic solutions to the previously mentioned problems, which come

from the real needs of the company in which this thesis is developed.

This document is organized as follows:

- Chapter 2 explains the use of 3D reconstruction as a tool of communication between the surgeon and the patient and plastic surgery planning. The second part of the chapter places the reader in the framework of this thesis, which has been developed in Crisalix. The readers will find a general overview of the framework used to obtain the 3d reconstruction and a description of its applications in the production company environment.

- Chapter 3 presents a review of the main approaches to perform 3D reconstruction. This chapter reviews the related methods of the state of the art to perform 3D reconstruction with special interest in face and breast 3D reconstruction.

- Chapter 4 presents a method to perform face alignment in images. This method is based on introducing prior knowledge, such as previously detected landmarks, into the shape statistical model in order to apply shape constraints. The method provides a weighted regularized projection into the parameter space. The outcomes of this research have been presented in the following paper:
  Guillermo Ruiz, Eduard Ramon, Jaime García, M. A. González, Federico Sukno. *Weighted regularized ASM for face alignment*. In proc. of the International Conference on Image Processing (ICIP), 2016.

- Chapter 5 presents a framework to fit a 3DMM of the breast to two possible inputs: 2D photos and 3D scans. The weighted regularized projection already introduced in chapter 4 is applied to 3D shapes. The contribution of each point in the 3DMM shape is weighted, allowing to assign more relevance to those points that we want to impose as constraints. The method is applied in several stages of the 3D reconstruction process. The publication associated to this contribution is:
  Guillermo Ruiz, Eduard Ramon, Jaime García, Federico Sukno,

M. A. González (2018). *Weighted Regularized statistical shape space projection for breast 3D model reconstruction*. Medical Image Analysis, 47:164-179.

- Chapter 6 presents the conclusions of this thesis and describes the possible lines of future work of the exposed research projects.

Chapter 2 have been removed from the thesis due to confidentiality issues.

CHAPTER 3

RELATED WORK

3D reconstruction of an object has been extensively studied in computer vision. Many reconstruction methods can be applied to recover the 3D structure of an object and especially the topic 3D face reconstruction has been addressed by the researchers is many publications. The chosen method depends on the scenario and the conditions in which the reconstruction is applied. When 3D reconstruction is applied to plastic surgery planning (*e.g.* face and breasts reconstructions) there are several requirements that need to be fulfilled. The quality of the reconstructed body parts needs to be suitable enough to be shown to a patient and the patient's anatomy must be represented very accurately. In addition, the input data for the method has to be easy to obtain.

In this chapter, we give an overview of the main approaches used for the researchers to perform 3D reconstruction. We have special interest in those methods focused on recovering the shape of human body parts such as the face or breasts. The reviewed 3D reconstruction methods can be classified depending on the scenarios and the needed input data. The possible scenarios are: **3D reconstruction from multiview images**, in which the input is a set of images describing the object from different points of view and **3D reconstruction from single image**.

3.1 3D reconstruction from multi-view images

3D reconstruction methods from multi-view images are widely used to recover the 3D shape of an object. The idea is to acquire multiple images of the object from different points of view, in order to estimate the surface of the object. Nowadays, 2D images can be easily obtained using common cameras which are accessible to everyone. These reconstruction methods can be classified depending on the scenario and the conditions in which the 2D images are captured. In some approaches, these conditions are controlled, meaning for example, that the set of cameras that acquire the images are calibrated or the lighting of the scene is known. These methods are described in section 3.1.1. On the other hand, there are methods in which all this scene information is unknown and the images are obtained under uncontrolled conditions. These methods are described in section 3.1.2.

3.1.1 Controlled scene

In this section, some of the most widely used methods to perform 3D reconstruction under controlled scenario are explained. For the particular case of reconstructing faces or parts of the human body, the most popular methods are multiview stereo, structured light or photometric stereo.

Multiview stereo is a well-known and widely used technique, in which the goal consists of recovering the 3D shape of an object from a set of images taken from different known positions [Seitz et al., 2006]. The problem of reconstructing an object can be solved either by passive or active methods. Passive methods are based on the acquisition of these images with fixed cameras usually with lighting setups. In this case, the only purpose of the lights is to illuminate the scene but without direct impact in the triangulation to estimate 3D points. The 3D points estimation is achieved by finding correspondences (*i.e.*, looking for the same point in two different images as illustrated in Figure 3.1) within the acquired images by using stereo matching methods. Using these multi-view setups for 3D reconstruction provides accurate results but also implies controlled

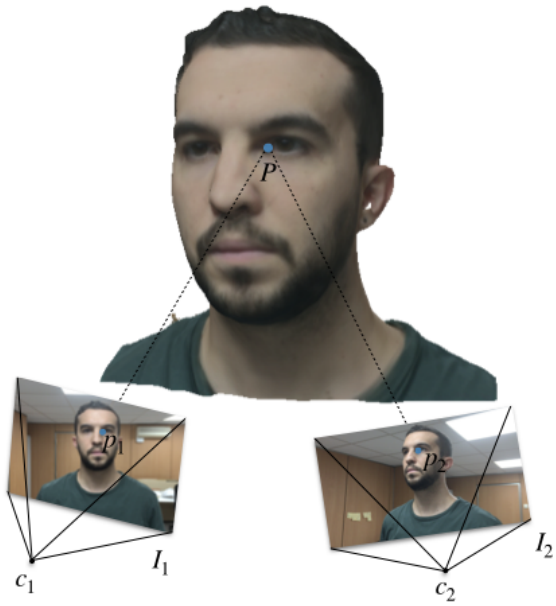


Figure 3.1: Stereo passive. Multiple images describing the reconstructed object are acquired from different points of view. In this example, two different cameras c_1 and c_2 are used to take the images I_1 and I_2 respectively. The blue dots p_1 and $p2$ represent the inner left eye bodymark in each one of the images. The point P (which corresponds to the left eye bodymark in the 3D object) can be estimated by triangulation using the points correspondences and the cameras.

scene conditions. In the particular case of reconstructing parts of the body the multi-view stereo setups typically consist in a set of calibrated cameras with illumination placed around the face [Beeler et al., 2010, Bradley et al., 2010, Beeler et al., 2011, Fyffe et al., 2014] or the breasts [Henseler et al., 2011, Hoeffelin et al., 2014].

Active stereo methods use additional information in order to obtain higher quality results and maximize their accuracy. This additional information usually consists of marker data placed in the scanned subjects [Alexander et al., 2009] or the use of structured light (see Figure 3.2). The

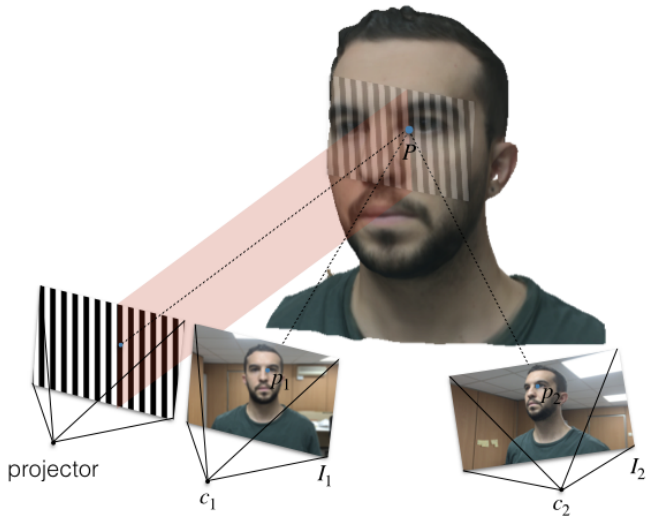


Figure 3.2: Stereo active. Multiple images I_1 and I_2 describing the scene are acquired from different points of view c_1 and c_2. Additionally, a device projects a pattern over the object (*i.e.*, projector). With this extra information the process of finding the correspondences p_1 and p_2 shown in blue dots is simplified.

process of finding correspondences is simplified and more robust with this additional information.

Structured light is a technology that consists in the use of a device projecting a light pattern (usually a simple geometric structures such as stripes, dots or grids) over the scene [Salvi et al., 2010] (see Figure 3.3). Another calibrated camera obtains the color image of the scene from a different point of view from the projected pattern. The knowledge of the pattern geometry and the recovered deformed patterns from the scene are used to extract the 3D surface information. The object can be scanned from different view points and the 3D surface information is integrated in dense 3D surface that recovers the scene depth. Structured light approaches are widely used in practical applications given their robustness. However, their accuracy directly depends on the resolution of the pro-

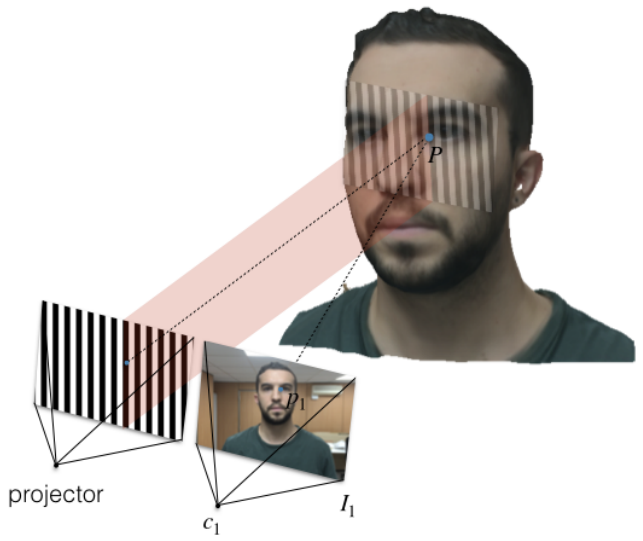


Figure 3.3: Structured light. The camera c_1 obtains the color image of the scene I_1. Another device projects a pattern over the scene (*i.e.*, projector). Knowing the pattern geometry and the suffered deformation of the projected pattern the depth of point P can be estimated.

jected patterns.

Photometric stereo methods were introduced by [Woodham, 1980] and address the problem of 3D reconstruction by estimating the 3D geometry of an object from multiple images of a specific sample obtained with a fixed camera and under varying lighting conditions. Computing the intensity variation in each pixel, the local orientation of the surface can be estimated. In addition, the reflectance albedo can be obtained as part of the same process. In order to simplify the computation, photometric stereo usually uses Lambertian reflectance model. The general equation of photometric stereo is formulated as:

$$\begin{bmatrix} I_{1,i} \\ I_{2,i} \\ I_{3,i} \end{bmatrix} = \rho_i \begin{bmatrix} L_1^T \\ L_2^T \\ L_3^T \end{bmatrix} n_i \tag{3.1}$$

for the specific case of having three different images with their respective light source vectors, where N is the number of the image pixels, $I_{k,i}; i = 1, ..., N$ is the set of intensities for image k and L_j is the jth light source vector. The reflectance albedo ρ_i and the surface normal n_i can be calculated by solving equation 3.1. However, the classic formulation of the method has some limitations: the images have to be acquired from the same pose under changing illumination conditions, which implies a specific capture device or setups. In addition, since all the images are taken from a single point of view, the estimated geometry is a partial region of the full object making difficult the reconstruction of the full object or a closed 3d surface. Usually, photometric stereo devices are used to capture the input images [Hansen et al., 2010, Angelopoulou and Petrou, 2014, Sun et al., 2015]. Some approaches combined multiview stereo or shape from motion with photometric stereo methods [Zhang et al., 2003, Esteban et al., 2008] dealing with images from different points of view and enabling a full 3D reconstruction.

The methods described in this section are able to obtain accurate 3D reconstructions but are subject to have some knowledge of the scene (*i.e.*, known cameras or illumination) or the use of specific devices that provide extra information. Other methods, such as photometric stereo, are usually too computationally expensive to be used in real time applications and they require training sets with collection of images with different illuminations of a specific sample.

3.1.2 Uncontrolled scene

More recent approaches perform 3D reconstruction from a set of images acquired in an uncontrolled way (*i.e.*, in the wild images) and are able to estimate the camera poses and the lighting. These methods are based in the aforementioned photometric methods. The reconstruction is performed using as an input a collection of images from an individual in which, there are a great variety of poses, expressions and lighting conditions without restrictions [Kemelmacher-Shlizerman and Seitz, 2011, Roth et al., 2015, Roth et al., 2016]. The authors of these works,

relax the constraints that are needed in photometric stereo (*i.e.*, rigid geometry, known lighting and cameras) to be able to work with collections of images acquired under unconstrained conditions. In the method of [Kemelmacher-Shlizerman and Seitz, 2011], the authors assume weak perspective camera and Lambertian reflectance. They detect fiducial points on the 2D images collection and use the corresponding positions of these fiducials in a 3D template to recover the pose of each face and normalize it (*e.g.* this results in a collection of normalized faces with all the faces in a frontal position). From this set of frontalized faces, the lighting and the surface normals are estimated. Similarly, the authors of [Roth et al., 2015] assume also weak perspective camera projection and Lambertian reflection. They perform 2D face alignment in all the images collection and a predefined 3D face template is deformed so that the projections of its 3D landmarks are consistent with the 2D detected landmarks. This deformation provides the global shape, but the fine details are obtained by estimating the photometric normals based on [Kemelmacher-Shlizerman and Seitz, 2011]. In the work of [Roth et al., 2016] the authors fit a 3DMM to produce a person specific template mesh instead of using predefined templates without prior face distribution knowledge. In addition, the contribution of each one of the images is weighted to use the ones that are more confident to produce accurate reconstructions in specific parts of the face. The authors of [Zeng et al., 2017] in order to avoid the use of generic templates or 3DMM, use multiple reference models, each one for some specific component of the input 2D face image (*e.g.* the eye, the nose or the mouth). With this combination, they claim that their method can obtain more accurate reconstructions.

In conclusion, these methods can recover fine details but they are usually too computationally expensive to be used in real time and they require a collection of images with different illuminations of a specific sample. Furthermore, to solve the challenging problem of performing reconstruction from images acquired in an unconstrained way, the authors usually need to make assumptions and introduce in their methods prior information such as predefined templates or 3DMM.

3.2 3D reconstruction from single image methods

In many applications the only available input is a single image of the subject that we want to reconstruct. This fact implies a great simplification of the input data (*i.e.*, a single image of the subject that can be easily obtained with a camera device) without the need of acquisition set ups, but nevertheless makes more challenging the 3D reconstruction problem. 3D reconstruction from a single image has been extensively explored by researches and is an ill-posed problem due to the difficulty to extract all the necessary information to perform the 3D reconstruction. Therefore, the methods that perform reconstruction from a single image, mitigate the uncertainty by introducing additional data (*i.e.*, prior information) such as illumination priors or statistical models. In this approach, common 3D reconstruction methods can be classified according to the introduced prior information in shape from shading, learning based methods and 3DMM models.

3.2.1 Shape from shading

Shape from shading methods are based on estimating the geometry of an object based on illumination and lighting changes applied to that object. The concept of shape from shading was firstly introduced in [Horn, 1975, Horn and Brooks, 1989] and since then, many different approaches appeared [Zhang et al., 1999]. Shape from shading methods estimate the surface normals of an object and are a particular case of the aforementioned photometric stereo methods. In this case, only a single image of the subject is available, making the ambiguity more extreme. The main principle of this technique is that the amount of light reflected by a surface is dependent on the orientation of the surface in relation to the light source and the observer. Therefore, these methods require knowledge about the illumination or lighting configuration, the surface reflectance properties, the color (*e.g.* albedo), the boundary conditions (*e.g.* depth knowledge in certain regions of the object) or the camera projection. However, all this

information is usually not available since it is part of the reconstructed object itself. Thus, these methods need to do assumptions and introduce certain constraints such as assuming that the lighting configuration is known or a few normals are previously given (*e.g.* the shape silhouette).

In particular, shape from shading has been extensively applied to perform 3D face reconstruction. Faces are difficult objects to reconstruct thus, in order to obtain suitable results for face reconstructions, authors have explored different approaches to introduce prior assumptions to reduce the problem difficulty. For example, the symmetry of the face has been used as a prior knowledge constraint to reduce the number of unknowns [Shimshoni et al., 2000, Zhao and Chellappa, 2000, Zhao and Chellappa, 2001]. Another approach to perform shape from shading is introducing statistical knowledge about the object. The first shape from shading statistical method was firstly introduced by [Atick et al., 1996] in which the authors represented the faces in a low-dimensional parameterization using principle analysis component (PCA). A different method that embedded statistical information in a shape from shading algorithm was [Smith and Hancock, 2006] in which the facial shape was described as the variation of surface normals directions. In the method of [Dovgard and Basri, 2004] the authors presented a combined approach which used faces symmetry and statistics of human faces as constraints in a restricted setup. Statistical methods are usually robust but they require a collection of instances with full correspondence in order to build the statistical model. In order to avoid this, other shape from shading approaches introduce a template model as reference for the reconstruction to supply the missing information and recover the shape information by estimating the rest of unknowns (*e.g.* lighting and reflectance) [Kemelmacher and Basri, 2006, Kemelmacher-Shlizerman and Basri, 2011, Jiang et al., 2018].

In conclusion, shape from shading methods can recover shape information and details from a single image but they require to take strong assumptions and usually the introduction of prior knowledge is needed. For face reconstruction, templates or statistical models are frequently used for these methods and the results are normally highly conditioned by them.

3.2.2 Learning-based methods

Learning-based methods relate the input image with some form of regression with the 3D reconstruction. These methods can be divided in regression methods and deep learning methods.

Regression methods have been extensively applied for the task of 2D face alignment [Xiong and De la Torre, 2013, Kazemi and Sullivan, 2014]. Given the good results of applying regression methods to 2D face alignment, the researchers have also started to use regression methods for 3D face reconstruction as well. Some approaches perform regression from a set of 2D landmarks to directly infer the 3D geometry of the face [Aldrian and Smith, 2010, Dou et al., 2014, Liu et al., 2015]. These methods consist on a pre-trained set of regressors that gradually adjust the 3D shape to an input image. The regressors are usually trained with a collection of paired 2D shapes (*i.e.*, sparse set of 2D landmarks) and their corresponding 3D shapes (with a known 2D to 3D shapes mapping). Thus, the regressors learn the existing correlation between them. Let P be defined as the annotated facial 2D landmarks $P = (x_0, y_0, x_1, y_1, ..., x_n, y_n)^t$ on a given image in which the landmark (x_i, y_i) as the ith landmark on the 2D image and n is the number of 2D landmarks. In order to reconstruct the 3D shape, the method is initialized with a the 3D shape S_0. The idea is to compute the adjustment that is applied to the shape S_{k-1} at the k_{th} iteration. In order to do so, the corresponding points to the landmarks are extracted from S_{k-1} and projected to the 2D image plane. In each iteration the landmarks displacement is defined as

$$\Delta P_{k-1} = P - \hat{P}_{k-1} \tag{3.2}$$

where $\hat{P}_{k-1}$ is the vector of the projected landmarks extracted from S_{k-1}. Then, the shape adjustment obtained using the regressor R_k is defined as

$$\Delta S_{k-1} = R_k(\Delta P_{k-1}) \tag{3.3}$$

The reconstructed 3D shape can be obtained in each iteration using

$$S_k = S_{k-1} + \Delta S_{k-1} \tag{3.4}$$

Once all the k regressors have been applied using equation 3.4 the result is the reconstructed 3D shape.

In the method of [Liu et al., 2016] the authors combine landmark regressors (*i.e.*, 2D points regressor) with shape regressors (*i.e.*, 3D points regressor). They alternatively use the landmarks regressor to locate the 2D points in the input image and then estimate the 3D shape with the shape regressor. At this point, the 2D regressed points are refined using the 3D shape. Thus, in each iteration there are three main steps: updating 2D the bodymarks, updating the 3D face shape and refine the 2D bodymarks. This regression approach results in a simultaneous face alignment and 3D face reconstruction method. Other methods perform regression from extracted features from the images [Jeni et al., 2015]. The regression algorithms proved to achieve good 3D face reconstructions but they are strongly dependent on the landmarks detection. In addition, specific details of the face can not always be recovered because the information to train the regressors is usually extracted from a set of sparse landmarks which do not cover the entire face.

Recently, deep learning approaches have successfully used networks to perform 3D object reconstruction. These methods are able to take advantage of all the image details and obtain accurate 3D reconstructions in milliseconds using convolutional neural networks (CNNs) from a single image. CNNs methods have shown the capacity of efficiently encode images into low dimensional vectors that express the variability of 3D data projected into the image plane. They can be divided in two main categories: model-based methods and model-free methods.

Given the difficulty of reconstructing an object from a single image, model-based methods introduce prior information about the reconstructed object. The authors usually make use of 3DMM in order to model plausible geometries, and train the CNNs to learn how to predict the shape parameters of the 3DMM given an input image [Richardson et al., 2016, Jourabloo and Liu, 2016, Zhu et al., 2016, Dou et al., 2017, Richardson et al., 2017, Tewari et al., 2017, Tran et al., 2017]. For example, the authors of [Richardson et al., 2016, Richardson et al., 2017] use synthetic data to train their network by artificially generating a large

face dataset images in correspondence with a 3DMM face representation. Later, their network learns how to predict the 3DMM parameters of a new face. While in [Richardson et al., 2016] the CNN is trained with synthetical data and an optional step of shape refinement with a shape from shading method is proposed, in [Richardson et al., 2017] the authors train two different newtorks. The first network is trained with synthetic data and recovers a rough estimation of the face geometry. The second network is trained with real images which recover the fine face details. In the method of [Jourabloo and Liu, 2016] the authors learn the 3DMM parameters and the projection vector that gives the relationship between 2D and 3D points. In this approach the authors use a cascade of CNN-based regressors, that perform their prediction in six stages using a different CNN in each one.

The prior information introduced by the 3DMM make the aforementioned methods robust in which the CNN learn 3DMM parameters which represent suitable face shapes. However, its use is subject to some limitations such as the availability of the 3D data to build the 3DMM or the limited expressivity of the 3DMM. For these reasons, some authors propose model-free methods, which avoid the use of 3DMM and present a more generic approach to reconstruct 3D objects. For example, the authors of [Sela et al., 2017] instead of using a method were the solution space is restricted to a low dimensional subspace they regress depth values per each pixel in the input image, representing the solution in a higher dimensional image domain. As a result, the network predicts two components: a depth image with the face geometry and a correspondence map which relates the image pixels and a facial template vertices. After these components are predicted by the network the authors perform a non-rigid registration of the template towards the depth data. As a result, the template with known parameterization adopts the geometry of the depth structure predicted by the network. Other methods propose to train the networks so they produce as an output a volumetric representation of the reconstructed object [Yan et al., 2016, Jackson et al., 2017, Häne et al., 2017]. These methods represent the reconstructed objects as dense voxel grids and train the networks with these volumetric representations. Although volumetric methods can

reach state of the art results, the main limitation that they present is that the volumetric grids of higher resolution are computationally demanding and the memory requirements are elevated. For this reason, authors try to minimize this restriction by optimizing the volumetric representation as done by [Häne et al., 2017] where the authors hierarchically predict small blocks of voxels in a coarse to fine resolution using octrees.

As a conclusion, CNN methods have successfully accomplished the task of reconstruction a 3D object. The prediction is usually fast once the network is trained and the predictions are accurate since the regressions are performed extracting details from the entire image and not only using sparse points. However, these methods require huge amount of training data to feed the neural networks and reach good results. In addition, there is a dependency on 3DMM since authors frequently use them in order to make the networks learn suitable representations of the modeled objects.

3.2.3 3D Morphable Models

3D Morphable Models (3DMM) were introduced by [Blanz and Vetter, 1999]. This method consists of obtaining a parametric description of the object and learning the space of allowable instances. This is done by building a statistical model of the object which basically consists of extracting the mean shape of the object and a number of variation modes from a set of training samples. The most common way to represent an object, such as a face or a breast shape, is with point distribution models (PDM). Each 3D shape x is formed by a set of n points defined as the vector $\boldsymbol{x} = (X_0, Y_0, Z_0, X_1, Y_1, Z_1, .., X_{n-1}, Y_{n-1}, Z_{n-1})^T \in \mathbb{R}^{3n}$ which contains the X, Y, Z coordinate of each point of the shape. The next step is to align all the training samples into the same coordinate frame. The most common way to do the shapes alignment is to apply generalized Procrustes Analysis [Gower, 1975]. Once all the training set shapes are aligned, the objective is to reduce the dimensionality of the training set. In order to represent 3D shapes of faces or breasts the number of needed points is usually large, thus applying a dimensionality reduction method, the set of main variation modes of the training data can be obtained. This

is done by applying Principal Component Analysis (PCA) [Jolliffe, 1986]. As a result, each shape x can be represented as:

$$\boldsymbol{x} = \bar{\boldsymbol{x}} + \boldsymbol{\Phi b} \tag{3.5}$$

where $\bar{\boldsymbol{x}}$ is the mean shape, $\boldsymbol{\Phi}$ is the matrix containing the eigenvectors of the covariance matrix ordered by largest eigenvalue which represents the main modes of variation. The first t main modes can be retained in order to reduce the dimensionality of the problem. Each eigenvector is weighted by the vector $\boldsymbol{b}$ which represents the shape $\boldsymbol{x}$ in the shape parameter space. In order to reconstruct a new 3D shape using a 3DMM, the shape parameters b need to be estimated.

3DMM have been extensively applied for representing the face anatomy. The main disadvantage of using 3DMM is that generating the statistical model can be a time consuming and laborious task because of the need of training samples which usually involves carrying out the processes of 3D data acquisition, the 3D data labeling and the registration of a reference mesh to the obtained data. For this reason, some researchers have made available their 3DMM versions. The authors of [Paysan et al., 2009] built their 3DMM from a training set of 200 scans (*i.e.*, 100 male scans and 100 female scans). More recently, the Surrey Face Model was released in [Huber et al., 2016]. The authors presented a multi-resolution 3DMM and methods for basic pose and shape fitting. The model was built from a training set of 169 scans.

The breast anatomy has been also represented using 3DMM. In the method of [Cordier et al., 2007] the authors built a PCA model of the breast using 28 instances of scanned breasts. The breast shapes variation is modeled by the statistical model which is used to construct modeler to synthesize realistic and plausible breast shapes. In the work of [Gallo et al., 2009, Gallo et al., 2010] the authors obtained women 3D breast shapes from 46 nuclear magnetic resonances (NMR). Since the NMRs present a great amount of noise the authors clean this data and register a reference template to build the PCA.

Once the 3DMM has been successfully built, it can be used in fitting algorithms to reconstruct and obtain accurate and complete representa-

tions of objects. There are multiple methods that researchers have applied to estimate the 3DMM parameters that best describe an input image. Traditional methods use an analysis-by-synthesis approach [Blanz and Vetter, 1999, Blanz and Vetter, 2003]. These methods are based on the optimization of a custom cost function which minimizes the difference between the input image and the model appearance. Usually, many parameters are estimated at the same time (shape, texture, pose or illumination) in a joint optimization. However, these methods are usually very time consuming and the optimization might be trapped in a local minima. In the work of [Romdhani and Vetter, 2005] the authors propose a method called Multi-Features Fitting which consists of adding information of the image, such as edges or specular highlights. In this manner, the optimization is constrained and easier to minimize. Nevertheless, the method requires the input of a set of manually labeled landmarks. Another different approach which has been widely used is the landmark based fitting [Asthana et al., 2011, Aldrian and Smith, 2013, Huber et al., 2016]. These methods estimate the 3DMM parameters by establishing correspondences between 2D landmarks detected in the input images and 3D landmarks. The most used approach in these methods is to optimize the parameters sequentially rather than in a joint optimization. For example, in [Aldrian and Smith, 2013, Zhu et al., 2015a] the authors propose to perform a sequential optimization by first estimating the camera pose using the affine model and shape parameters and later estimate illumination using spherical harmonics. Recently the authors of [Hu et al., 2017] present a sequential method which proposed two separate processes: a geometric model fitting and a photometric model fitting and claim that they perform faster than other existing methods. A common factor in most of the aforementioned methods is that they share the use of 2D landmarks detection in the input image to establish 2D-3D correspondences. The 2D landmarks detection is an essential step that has an important impact in the following stages of these methods and therefore conditions the final result of the 3D reconstruction. For this reason, we start to work with 2D shapes and explore the introduction of prior information into ASM. With the goal of developing a robust 2D shape detection method we detail the proposed method in chapter 4.

Afterwards, in chapter 5 we extend the methodology to work with 3DMM and overcome some of the limitations previously mentioned.

The content of the following chapter is associated to the publication:

Guillermo Ruiz, Eduard Ramon, Jaime García, M. A. González, Federico Sukno. *Weighted regularized ASM for face alignment*. In proc. of the International Conference on Image Processing (ICIP), 2016.

CHAPTER 4

WEIGHTED REGULARIZED ASM FOR FACE ALIGNMENT

Active Shape Models are a powerful and well-known method to perform face alignment. In some applications it is common to have shape information available beforehand, such as previously detected landmarks. Introducing this prior knowledge to the statistical model may result of great advantage but it is challenging to maintain this priors unchanged once the statistical model constraints are applied. We propose a new weighted regularized projection into the parameter space which allows us to obtain shapes that at the same time fulfill the imposed shape constraints and are plausible according to the statistical model. The performed experiments show how using this projection better performance than competing state of the art methods is achieved.

[1]The content of this chapter is associated to the following publication:

Guillermo Ruiz, Eduard Ramon, Jaime García, M. A. González, Federico Sukno. *Weighted regularized ASM for face alignment*. In proc. of the International Conference on Image Processing (ICIP), 2016.

4.1 Introduction

Active Shape Models (ASM) are a widely used approach for object segmentation tasks and more specifically have been successfully used for face alignment applications. A statistical shape model is built from training examples, modeling the variation that the object class contains within all samples. The ASM uses this statistical information to impose shape constraints (*i.e.*, model constraints) in order to ensure obtaining valid shapes during the segmentation process. ASMs have proved to be a robust method, however some limitations exist. For example, their accuracy depends on the quality of the shape initialization, which shall be sufficiently good to prevent getting stuck in local minima. The existence of prior information (*e.g.* in some applications the location of a small set of points is available, because they either have been previously detected or provided by the user) is of vital importance and can be used to perform a better initialization of the ASM algorithm, and to estimate the rest of the shape from this small set of points (*i.e.*, initial constraints). The imposition of the model constraints makes difficult (or even impossible in many cases) the task of obtaining plausible shapes preserving the exact position of the original known points.

In this chapter, following the work of [Zhao et al., 2004], we present a weighted regularized projection into the parameter space which can be used to: 1) initialize shapes given a small set of known points, and 2) introduce this prior information during the search procedure by projecting the searching shape in a weighted manner and applying a regularization which favors probable shapes according to the statistical model. In this way, it is possible to obtain such shape parameters that maintain almost unchanged the position of the constrained points and represent in a realistic way the shape of the modeled object. The method has been evaluated in the public face database MUCT [Milborrow et al., 2010] and compared with another competing state of the art method, obtaining better results. The rest of the chapter is organized as follows: section 4.2 describes the related work on introducing constraints by adding prior information to an ASM. In section 4.3, we detail our weighted regularized ASM projection.

The experimental results are shown in section 4.4 and the conclusions are exposed in section 4.5.

4.2 Introducing constraints in ASM

In an ASM the object's shape is represented with point distribution models (PDM) which can be defined with a set of n landmark points $\mathbf{x} = (x_0, y_0, x_1, y_1, ..., x_{n-1}, y_{n-1})^T$. In order to build the statistical model, principal component analysis (PCA) is applied to the aligned set of training shapes. The principal modes of variation explain in which ways the points of the shape tend to move together and are described by the eigenvectors associated to the largest eigenvalues. Any new instance of the represented object x can be obtained then, as the sum of the mean shape $\bar{x}$ and a linear combination of the shape parameters with the following equation:

$$x = \bar{x} + \Phi b \tag{4.1}$$

where $\Phi = (\Phi_0, \Phi_1, ..., \Phi_k)$ is the matrix of the first t eigenvectors ordered by largest eigenvalue and $b = (b_0, b_1, ..., b_k)^T$ is a vector containing the shape parameters which weights the contribution of each eigenvector.

In order to ensure that the obtained shapes are valid, the ASM restricts the shape parameters space by enforcing model constraints. This is typically done by applying one of the common shape constraint strategies such as hyperrectangle projection [Cootes et al., 1995], scaling [Stegmann et al., 2000] or hyperelliptical approximation [Cerrolaza et al., 2011]. This correction modifies the predicted shape parameters, adjusting them to ensure the plausibility of the shapes. Knowing the correct positions of a small set of points leads, on the one hand, to a better and more robust initialization. In addition, during the search procedure (*i.e.*, fitting) these points can remain at their correct positions without the need of finding new positions for them. This means that we can rely on these known points to estimate the positions of the rest of the points of the shape. If the mentioned prior information is used together with the statistical model, it is possible to increase the performance of the model. The problem arises

when the ASM applies one of the aforementioned model constraint strategies. Such readjustment modifies the position of all the points of the shape causing displacement in the initial constraints and moving them from their correct position, which is not desired.

The approach of predicting a shape exploiting some available prior information has been previously explored and used in different kinds of applications. In this section, we describe some methods relevant to our work.

Some approaches weigh the contribution of certain points of the shape as seen in [Zhou et al., 2013] where the authors present a facial landmark localization method. The algorithm learns a weight vector for each landmark that minimizes the sum of reconstruction error and a regularization term that distributes the weights to other landmarks that could carry important information rather than obtaining a sparse weight vector. The optimal set of candidate points is found by solving a graph matching problem. In [Zhao et al., 2004] the authors propose to perform an evaluation of the fitting step in the ASM using local appearance models and with this information, find the shape parameters performing a weighted projection in the shape space. During the fitting procedure, the Mahalanobis distance between the current shape candidates and the local appearance model mean is minimized to find the best candidate points.

Other methods use a set of known points to infer the position of the rest of the points in the shape, as done by [Blanc et al., 2009]. Given a statistical model and a partially observed region of the shape the authors aim to use this information to predict the rest of the shape and establish a region of confidence around the predicted shape. They propose the following formula to compute the shape parameters:

$$\begin{aligned} b_\lambda &= \underset{b}{Argmin}(\parallel \Phi_x b - \mathring{x} \parallel + \lambda b^T \Lambda^{-1} b) = \\ &= (\Phi_x^T \Phi_x + \lambda \Lambda^{-1})^{-1} \Phi_x^T \mathring{x} \end{aligned} \tag{4.2}$$

where Φ_x corresponds to the decomposition of the eigenvectors matrix in sub matrices containing the vectors that affect the predictors and the rest of variables respectively, $\mathring{x}$ is the centered shape $(x - \bar{x})$, Λ is the matrix

of the retained non-zero eigenvalues of the covariance matrix and λ is an additional regularization term. The first term of the formulation is enforcing that the predicted parameters make the new shape fit the already observed points as closely as possible. The second term with the regularization λ indicates how probable our predicted shape is according to the statistical model. In the work of [Hug et al., 2000] an interactive framework for shape initialization is proposed, based on the interaction with the user who introduces a small set of points and those are used to compute the most probable shape containing the provided points. The idea is to iteratively remove the variation introduced by these initial points to the statistical model in order to maintain them at their correct position. This task is achieved by computing two basis vectors δ_x and δ_y which produce a unit displacement in the x- and y- direction respectively for the fixed constraints. Many vectors satisfy this property; however, the one that minimizes the Mahalanobis distance to the mean is chosen. This process is applied to every constrained point and a new PCA is computed each time in order to determine the shape model with the removed variation at the current fixed point. Finally, the algorithm provides the most probable shape according to the statistical model that fits all the given constrained points. Following this approach, in [van Ginneken et al., 2003] the authors present an interactive version of the algorithm where the user can drag a few points of the object's outline to their correct positions improving the accuracy of the ASM. Another application where it is possible to use a small set of known landmarks as constraints in order to automatically trespass additional landmarks from different dataset sources, can be found in [Smith and Zhang, 2014].

Another common strategy is to use a probabilistic approach in order to introduce prior information. In [Cootes and Taylor, 2001] the authors propose the Constrained Active Appearance Models reformulating the original AAM using a probabilistic framework. This method allows to combine the previous scheme with a user input, which sets constraints in some points of the shape. This leads to an improvement of the mean boundary position error; however as the number of constraints is increased, the texture error increases as well. A reconstruction method for an extracted

shape applied to vertebral fracture detection is proposed in [de Bruijne et al., 2007]. The idea of the authors is to compare an observed vertebral shape with its reconstruction to measure the variation and detect fractures or anomalies. In order to do so, the authors propose to model the variation over all the shape subjects and, in addition, model the variation between pairs of vertebrae in the same subject. The introduction of this extra prior information allows the model to distinguish between healthy vertebrae and deformed or fractured vertebrae. The variation over an individual subject is modeled with a PDM as usual, and conditional PDMs are used to reconstruct the most likely shape of a vertebra, given a different known vertebra in the image (prior information). Similar methods of the state of the art have been also applied to predict 3D shapes from available information. In [Rajamani et al., 2007], the authors aim to recover a 3D shape from a small subset of available points. Similarly to other methods explained above, the idea is to predict a new 3D shape from the linear combination of the mean shape and the shape parameters while trying to minimize the error between the predicted shape and the already known points. The objective function minimized is

$$f = \rho \left\{ \sum_{\substack{k=1 \\ j=Index^{(k)}}}^{N} w_k \parallel y_k - (\overline{x}_j + \sum_{i=1}^{m} b_i \Phi(j)) \parallel^2 \right\} + \\ + (1-\rho) \left\{ \sum_{i=1}^{m} \frac{b_i^2}{\Lambda_i} \right\} \tag{4.3}$$

where the first term of the formula is the Mahalanobis distance between the N available points y and the mean shape $\overline{x}$ plus a weighted sum of the eigenvectors. The second term introduces the Mahalanobis distance between the predicted shape and the mean which enforces that the predicted shapes are inside the distribution according to the model. The parameter ρ is to control how much we want to keep our shape similar to the mean shape or allow it to deform to better fit the control points. As a result of the proposed algorithm, 3D shapes can be estimated giving as an input

small or large set of points obtaining a realistic representation of the volume. Another example can be found in the work of $[Suknoet\ al., 2015]$. Sets of possible candidate locations are generated under the assumption that some points might not be accurate at the input set or might be unknown. Using combinatorial search and a statistical model, the method infers the position of these unavailable landmarks by predicting probable shapes. The objective is to infer the position of these landmarks maximizing the probability of the shape according to the model and ideally without changing the position of the known points.

4.3 Weighted Regularized ASM

We propose a weighted regularized ASM which can be used for fitting shapes, giving high weights to points that we can rely on and at the same time obtaining plausible shapes. We follow the work presented by [Zhao et al., 2004]. If we consider that x^p is the principal component projection of x, then the squared reconstruction error is:

$$ERR = \|x - x^p\|^2 = (x - x^p)^T(x - x^p) \tag{4.4}$$

Taking this into account, a point weighted projection is performed in the shape space in order to make the solution closer to the most probable points. The most relevant points (*i.e.*, the points which come from the local appearance model distribution and are close to the model mean) will have higher weights because they are more reliable since they are closer to the solution than the others, which introduces another reconstruction error:

$$ERR_W = (x - x^p)^T W(x - x^p) \tag{4.5}$$

Where $W$is a diagonal matrix of weights. Finally from minimizing ERR_W the weighted projection in the shape space is obtained as:

$$b = \left[\Phi^T W \Phi\right]^{-1} \Phi^T W(x - \bar{x}) \tag{4.6}$$

Zhao et al. claim that this weighted ASM gets more accurate results and it is more robust than the common ASM method. This projection is suitable

to our purpose, since we can apply high weights to the points that correspond to the constraints to make the predicted shape fit them and predict the rest of the points given this prior knowledge. However, to ensure that the predicted shape is plausible given the shape model, the shape parameters need to be constrained in the subspace of allowed shapes. We aim to find a projection able to maintain the position of the initial constraints and promoting probable shapes. We propose to introduce into the formulation the Mahalanobis distance to the mean as a regularization term that will favor shapes similar to the model's mean. To do so, we define two energies, E_D which is the energy corresponding to the data term extracted from equation 4.5 and E_R which corresponds to the regularization term:

$$\begin{aligned} E_D(b) &= (x - x^p)^T W (x - x^p) = \\ &= (x - \bar{x} - \phi b)^T W (x - \bar{x} - \phi b) \qquad (4.7) \\ E_R(b) &= b^T \Lambda^{-1} b \qquad (4.8) \end{aligned}$$

A general energy E is defined adding the two previous energies and introducing a parameter λ to tune the regularization term.

$$\begin{aligned} E(b) &= E_D(b) + \lambda E_R(b) \\ E(b) &= (x - \bar{x} - \phi b)^T W (x - \bar{x} - \phi b) + \lambda b^T \Lambda^{-1} b \qquad (4.9) \end{aligned}$$

Finally, we derive to find such b that minimizes the defined energy E:

$$\begin{aligned} \nabla_b(E(b)) \equiv \nabla_b(&-x^T W \phi b + \bar{x}^T W \phi b - (\phi b)^T W x \\ &+ (\phi b)^T W \bar{x} + (\phi b)^T W \phi b + \lambda b^T \Lambda^{-1} b) \qquad (4.10) \end{aligned}$$

Then, we finally get a closed form solution:

$$b = (\phi^T W \phi + \lambda \Lambda^{-1})^{-1} \phi^T W (x - \bar{x}) \qquad (4.11)$$

Using equation 4.11, we can perform a projection in the shape space weighting the contribution of each one of the points in the shape and additionally control the shapes plausibility with the regularization term λ. Higher values of the regularization term will favor shapes closer to the mean shape.

4.4 Experiments

The method has been tested on the widely known public MUCT database [Milborrow et al., 2010]. The faces in the MUCT database are labeled with 76 manual landmarks. For our experiments we have used the subset (a) in which the camera is located in front of the subject's face and is composed by 750 faces. As evaluation method we have performed cross-validation, choosing 10 as the number of folds. We have computed the distance between the fitted shapes with their corresponding manually labeled shapes. The measured point-to-point alignment errors have been normalized by the inter pupillary distance. In all the images the fitting is initialized using a subset of known points of variable length. In Figure 4.1 the accuracy of the algorithm is shown given a small set of 4 initial constrained points applying high weights on them. As observed, the weighted regularized ASM maintains their position. The error in these points is hardly visible while the highest errors are found in the outline of the jaw. We have also compared the accuracy of our weighted regularized ASM against another state of the art method which also fixes initial constraints proposed by Hug et al. [Hug et al., 2000]. Both algorithms initialize the mean shape into the image and estimate the rest of shape points using the initial constraints. During the fitting both approaches aim at maintaining the known points at their original position. The cumulative error distribution (CED) curves are shown in Figure 4.2. The results show how the performance of both methods is similar but using the weighted regularized ASM the mean point to point error is slightly lower and in all cases the worst case error is lower when using our projection. Another advantage is shown in Figure 4.3 evidencing that the obtained shapes are more plausible when using our approach thanks to the regularization. However, we also observed that, as the number of initial constraints is increased, the distributions in both methods tend to be more similar. In addition, our method provides a computational gain, since a unique statistical model is needed, as opposed to [Hug et al., 2000], where the statistical model has to be rebuilt each time that an ASM point is set as a constraint. Finally, in Figure 4.4 we show fitting examples compar-

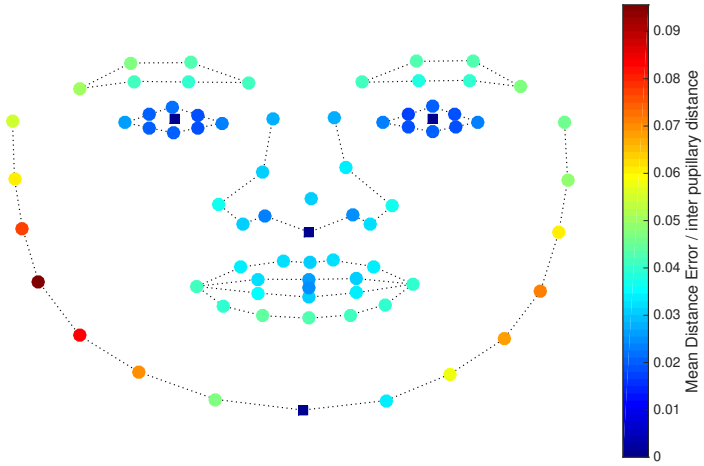


Figure 4.1: The point-to-point mean distance alignment error given a small set of 4 constrained points. Note that the landmarks with known location have square shape, with almost inexistent error since their position has been preserved.

ing [Hug et al., 2000] method and the weighted regularized ASM using the same set of initial constraints in both methods. As can be observed, both methods provide suitable shapes and preserve the location of the initial constraints. However, if we look closer to some details such as the detection of the eyebrows landmarks or the face contour the weighted regularized method performs better than the compared method thanks to the combination of regularization term and the weighted points.

4.5 Conclusions

In this chapter we have presented a weighted regularized projection which allows giving more importance to those points that are more reliable than others. At the same time our projection ensures obtaining plausible shapes according to the statistical model through a regularization term. This method can be used as initialization of the ASM and during the fitting in

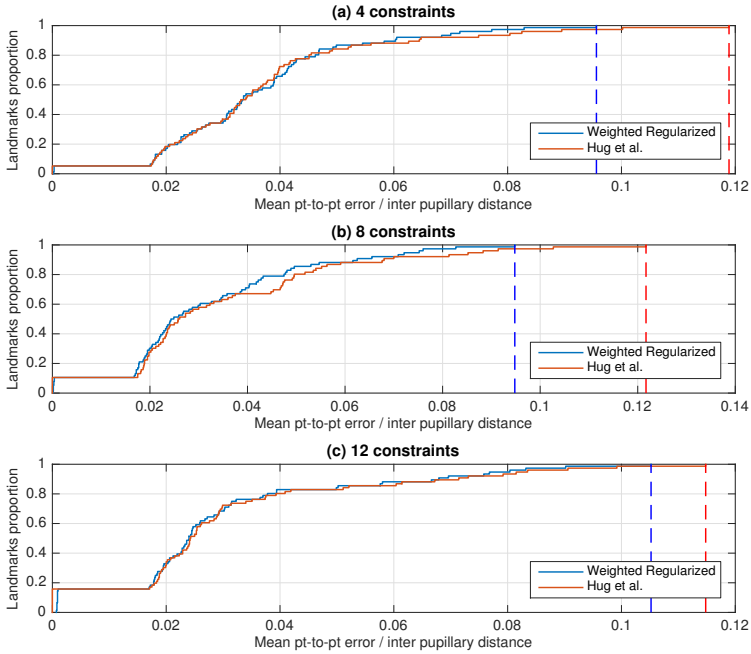


Figure 4.2: CED curve comparing the accuracy of our weighted regularized ASM with [Hug et al., 2000]. In (a), (b) and (c) the results are shown using a set of 4, 8 and 12 initial constraints respectively. The highest error point in each one of the curves is indicated with a discontinued line.

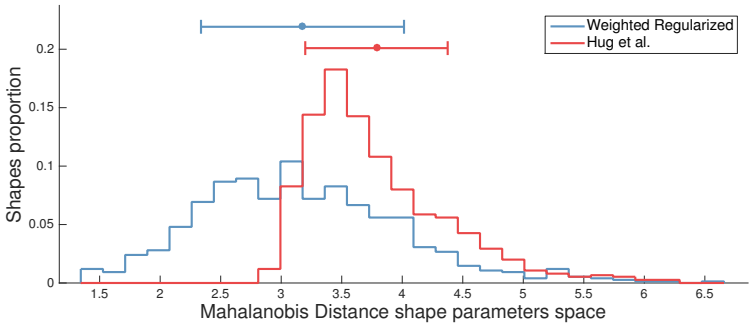


Figure 4.3: Histogram of the Mahalanobis distance to the mean in the shape parameters space. The lines on the top show the mean and the standard deviation of each distribution.

Figure 4.4: ASM fitting examples comparison. The first row shows the resulting fitted shapes using [Hug et al., 2000] method and the second row shows the resulting fitted shapes using the Weighted regularized ASM.The 76 MUCT landmarks are shown in colors: green points correspond to anatomical face bodymarks, blue points correspond to contour bodymarks and red points correspond to the set of bodymarks used as fixed initial constraints.

order to ensure the preservation of the position of a set of reliable points and guarantee the shapes plausibility. We have evaluated the method in a public database showing that it performs better than the most directly competing state of the art method.

The content of the following chapter is associated to the publication:

Guillermo Ruiz, Eduard Ramon, Jaime García, Federico Sukno, M. A. González.*Weighted Regularized statistical shape space projection for breast 3D model reconstruction*. Medical Image Analysis 2018, 47:164-179.

CHAPTER 5

WEIGHTED REGULARIZED STATISTICAL SHAPE SPACE PROJECTION FOR BREAST 3D MODEL RECONSTRUCTION

The use of 3D imaging has increased as a practical and useful tool for plastic and aesthetic surgery planning. Specifically, the possibility of representing the patient breast anatomy in a 3D shape and simulate aesthetic or plastic procedures is a great tool for communication between surgeon and patient during surgery planning. For the purpose of obtaining the specific 3D model of the breast of a patient, model-based reconstruction methods can be used. In particular, 3D Morphable Models (3DMM) are a robust and widely used method to perform 3D reconstruction. However, if additional prior information (*i.e.*, known landmarks) is combined with

[1]The content of this chapter is associated to the following publication:

Guillermo Ruiz, Eduard Ramon, Jaime García, Federico Sukno, M. A. González (2018).*Weighted Regularized statistical shape space projection for breast 3D model reconstruction*. Medical Image Analysis, 47:164-179.

the 3DMM statistical model, shape constraints can be imposed to improve the 3DMM fitting accuracy.

In this chapter we present a framework to fit a 3DMM of the breast to two possible inputs: 2D photos and 3D point clouds (scans). Our method consists in a Weighted Regularized (WR) projection into the shape space. The contribution of each point in the 3DMM shape is weighted allowing to assign more relevance to those points that we want to impose as constraints. Our method is applied at multiple stages of the 3D reconstruction process. Firstly, it can be used to obtain a 3DMM initialization from a sparse set of 3D points. Additionally, we embed our method in the 3DMM fitting process in which more reliable or already known 3D points or regions of points, can be weighted in order to preserve their shape information. The proposed method has been tested in two different input settings: scans and 2D pictures assessing both reconstruction frameworks with very positive results.

5.1 Introduction

During last years, 3D reconstruction has been of great interest for researchers due to the great variety of applications in which it can be used such as medical image analysis, video games, film industry, virtual reality or animation. The literature about the different areas in computer vision that are involved in 3D reconstruction has been considerably increased. One of the fields where this technology can be applied and where 3D reconstruction has gained popularity is in the plastic and reconstructive surgery sector. Physicians are becoming more and more interested in introducing 3D reconstruction in their professional activities since it offers a non-invasive procedure to study the patient's anatomy. Moreover, it can help improving the communication between patient and surgeon during the surgery planning process allowing to simulate the possible outcomes of the surgery in a realistic way. Nevertheless, the usual approaches used to assess patients are based in 2D photos which might not reach a sufficient detail level. Other systems such as scanners or magnetic resonance

imaging which are able to reproduce the body anatomy very accurately, happen to be very time consuming, invasive and too costly to be introduced in the medical daily consultation.

The aim of this work is to obtain 3D reconstructions representing accurately the breast anatomy of plastic surgery patients from a vision-based system not requiring specialized equipment. Once the patient's 3D model is obtained, it can be used to simulate any breast surgery procedure by modifying its shape.

5.1.1 3D model-based reconstruction

Model-based methods are a generic approach to perform 3D reconstruction. They rely on introducing a prior model with the shape information of the reconstructed object. This model is fitted to the new instances preserving the structure and shape information. Specifically, statistical models built from a set of training shapes, have proven to be a robust model-based method since statistical restrictions can be imposed and therefore obtain suitable reconstructions while avoiding unlikely shapes.

3D morphable models (3DMM) [Blanz and Vetter, 1999] are a powerful tool and have been already used by some authors with the purpose of 3D breast reconstruction [Cordier et al., 2007, Kim et al., 2008]. 3DMM consist of a statistical shape model obtained by applying Principal Component Analysis (PCA) to a set of training shapes. Thus, new instances can be efficiently represented by a linear combination of the model parameters. In order to obtain the 3D reconstruction the 3DMM is fitted to a target shape. This is achieved by searching correspondences between the target shape and the 3DMM and finding such shape parameters that best represent the target shape given the 3DMM model. Additionally 3DMM can also include a texture model which is built separately from the shape model.

Although the use of 3DMM presents great advantages some limitations have to be taken into account. A good initialization of the 3DMM fitting process is vital for a proper reconstruction. Several aspects depend on the initialized shape such as finding correct correspondences with re-

spect to the target shape or the convergence of the fitting process. Usually a predefined mesh template is used for initialization purposes, which sometimes might be too distant from the target shape. The prediction of a 3D shape introducing prior information into a statistical model has been used in order to improve the mesh initialization [Hug et al., 2000, Rajamani et al., 2004] or to obtain 3D reconstructions from sparse data [Sierra et al., 2006, Zheng et al., 2006, Rajamani et al., 2007].

Another problem arises when the statistical restrictions are applied in order to preserve shape plausibility. The knowledge provided by the 3DMM allows to identify the shapes that belong or not to the subspace of allowed shapes. It is possible to approximate a shape to this subspace to ensure its plausibility [Cootes et al., 1995, Stegmann et al., 2000, Cerrolaza et al., 2011]. However, due to this approximation, all the shape is modified while sometimes it would be desirable to preserve the position of certain landmarks previously detected or whose positions are known beforehand.

5.1.2 Contribution

Our main contribution is a Weighted Regularized (WR) projection, which can be used to solve the aforementioned problems. The main idea of the WR projection is to combine the 3DMM statistical shape model and additional prior information, such as sparse set of known landmarks. In this way, plausible shapes satisfying additional shape constraints can be obtained in one single step. Given a small set of landmarks, it is possible to initialize a 3D shape that is statistically plausible and fulfills the position of these points. This provides an initial shape predicted by the model with certain constrained points which makes it closer to the target shape than a predefined template. Furthermore, we embed our WR projection into the fitting process based in a non-rigid Iterative Closest Point (ICP). Those landmarks or regions of points that we can rely on, can be weighted in order to preserve their shape information when projecting a shape into the parameter space. Our results show that introducing the proposed WR projection in the 3D reconstruction process improves both the initialization

and fitting processes.

The presented framework is capable to fit the 3DMM to two different inputs: 3D scans (*e.g.* obtained from a structured light scanner) and multiple 2D photos from the patient taken from different viewpoints. The scheme of the main framework components is shown in Figure 5.1.

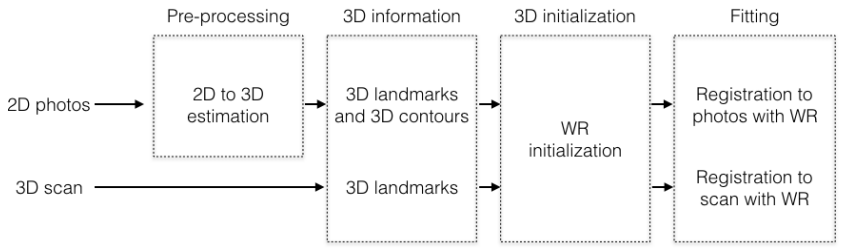


Figure 5.1: Scheme of the proposed 3D reconstruction framework. Two possible workflows are considered depending of the input information, either 2D photos or a 3D scan. A pre-processing step is needed when using 2D photos in order to estimate the detected landmarks and contours in 3D space. We then use our WR projection to obtain an initial mesh estimation in both workflows. Starting from this initialization the fitting process is performed registering the 3DMM towards the available information in each case (*i.e.*, the 3D estimated contours or the scan, respectively). In the fitting step, the WR projection is also used to preserve the plausibility of the deformed meshes.

This work is built on top of a previous conference publication [Ruiz et al., 2016] in which a preliminary version of our WR projection was presented in the context of 2D face alignment. This chapter extends the work previously done for face alignment in 2D photos to shapes in 3D space. The contributions of this work are:

- A framework to fit a 3D Morphable Model representing the breast anatomy of the patient using either 3D scans or 2D photos.
- A weighted regularized projection in the parameter space which

allows to obtain plausible shapes according to a statistical model while additional constraints can be imposed to preserve the position of known landmarks.

- A reformulation of the method [Ruiz et al., 2016] allowing to regularize a given shape towards a prior shape (*i.e.*, a known shape or a template), which is not necessarily the mean shape.

The rest of the chapter is organized as follows: we review the interest in introducing 3D imaging in plastic surgery planing, 3D model based reconstruction and the related work on introducing constraints in statistical models in Section 5.2. We detail our WR projection method and describe the 3DMM fitting framework for scans and photos in Section 5.3. We then provide experimental results of 3D reconstruction from both frameworks, measuring the impact of introducing the WR projection in the reconstruction process in Section 5.4. Finally, we present the conclusions in Section 5.5.

5.2 Related work

5.2.1 Breast surgery planning

3D imaging is becoming a popular tool in breast surgery planning [Chang et al., 2015, O'Connell et al., 2015] and is gaining acceptance among plastic and reconstructive surgeons. It presents a non-invasive and safe method to represent the patient's breast anatomy. In plastic surgery the patient's opinion and thoughts before the surgery process are very important. This can be supported by a realistic visual representation of their body and a visualization of the possible final result. Using 3D imaging it is possible to generate a virtual visualization so that patients can see their own body in terms of 3D models and furthermore, once the 3D model is obtained, patients would be able to preview the simulation of plastic or aesthetic procedures by modifying the 3D model. This presents a great advantage in terms of visualizing and planning plastic surgery or surgical

interventions in a non-invasive way. There are two main approaches that have been used to obtain the 3D reconstruction of the breast:

- Stereo vision: Using as an input 2D images of the same scene taken from different positions [Henseler et al., 2011, de Heras Ciechomski et al., 2012, Hoeffelin et al., 2014].
- 3D scanners: 3D scans are obtained using specific hardware, providing an accurate result of the patient's shape and texture [Tzou et al., 2014]. This kind of equipment can be very expensive and with big dimensions which complicates its installation at the doctor's offices. Such scanning hardware is provided by companies like Axisthree™ or Canfield™.

The 3D reconstruction of the breast can have different applications. Some methods perform assessment of the breast shape or volume extracting metrics or useful information such as the breast symmetry or size [Kovacs et al., 2006, Cordier et al., 2007, Chen et al., 2010, Veitch et al., 2012]. Other methods, simulate plastic or aesthetic procedures to advise the patients during the pre-operative process and show the possible outcome of their surgery [de Heras Ciechomski et al., 2012, Georgii et al., 2014].

In this chapter we propose a model-based method to perform the 3D reconstruction of the breast with a framework for each one of the mentioned 3D reconstruction approaches. The input for our method can be scans obtained with a 3D scanner or standard 2D pictures.

5.2.2 3D model-based reconstruction

Model-based reconstruction is a popular approach to determine the geometry of objects in 3D. It consists on the incorporation of prior knowledge about the reconstructed object by using a generic template or statistical model obtained from a training database. This method is robust to deal with problems such as noisy data, incomplete scans or occluded elements since the model allows to complete a full reconstructed object according to the learning data and with known parameterization.

In the context of facial gesture recognition some early approaches used a generic template to perform 3D reconstruction without using statistical model avoiding the need of a training set to build the model which can be very laborious to obtain. [Liu et al., 2001] instead of representing a shape with a linear combination of the statistical model parameters, represented a face as a linear combination of a pre-designed template and different vectors that linearly deform the different parts of the face. [Pighin et al., 2006] used a scattered data interpolation technique to deform a generic template to fit the input geometry to recover different facial expressions. These approaches avoid the generation of statistical models but once the generic template is deformed the resulting shapes plausibility is not ensured.

3D model-based reconstruction using statistical models is a recurrent and widely used method due to its robustness. 3DMM were introduced by [Blanz and Vetter, 1999]. They are usually composed by a shape model and a texture model. The shape is represented by a shape vector $\boldsymbol{S} = (X_1, Y_1, Z_1, X_2, \ldots, Y_n, Z_n)^T \in \mathbb{R}^{3n}$, which contains the 3D coordinates X,Y,Z of each point of the n shape vertices. The texture is represented with a texture-vector $\boldsymbol{T} = (R_1, G_1, B_1, R_2, \ldots, G_n, B_n)^T \in \mathbb{R}^{3n}$, which contains the R,G,B color values for the n vertices. PCA is applied to both shape vector and texture vector. Once the models are built, any new shape instance can be expressed as a linear combination of the shape vector and the texture vector. Some authors have already used 3DMM to represent the breast anatomy [Cordier et al., 2007, Kim et al., 2008] and statistically model the breast shape.

3DMM can be used to reconstruct a shape having as an input 2D images or 3D depth information. The most used approach to reconstruct a 3D shape from 2D images using 3DMM is analysis-by-synthesis [Blanz and Vetter, 1999, Blanz and Vetter, 2003]. Shape parameters are found by minimizing the error between preestablished correspondences between the 3DMM and input image points. On the other hand, in the last years, regression based methods [Huber et al., 2015, Zhu et al., 2015b, Jourabloo and Liu, 2016] have obtained great results. These methods extract features around detected key points in the images and perform regression.

3DMM are also used when depth information (*e.g.* 3D point cloud or scan) is available. In this case, problems arise such as incomplete scans, holes, noise or unknown mesh parameterization. To solve these issues, the most common approach is to fit the 3DMM towards the scan (*i.e.*, registration). The iterative closest point (ICP) algorithm [Besl and McKay, 1992] and its variants are the most widely used approaches. These methods consist in looking for correspondences by searching the closest points between the source and the target mesh, and finding the transformation that aligns these correspondences [Brunton et al., 2014, Zhu et al., 2015b].

More examples of 3DMM applications and different uses can be found in [Kittler et al., 2016, Jourabloo and Liu, 2017].

5.2.3 Introducing constraints in 3D Shape Models

Although 3DMM is a robust method for the purpose of 3D reconstruction some limitations exist. Firstly, an initial shape is needed to initialize the fitting process, which in some cases can be distant from the target mesh. The most commonly used approach to initialize the fitting process is using a predefined template or the mean shape of the statistical model. This is sufficient for instances which are not very far from the predefined template. However, if the target shape is distant, this can lead to finding incorrect correspondences or to the non convergence of the fitting process.

In addition, during the fitting process the plausibility of the shapes needs to be ensured. The statistical model allows to approximate a given shape to the subspace of allowed shapes. Due to this approximation, the shape parameters are modified getting a new corrected shape as output (see Section 5.3.1 for more details of how this shape approximation is performed). As a result, the expectation with respect to the shape model is maximized, but not the likelihood towards the original shape. Thus, the positions of previously known landmarks that we might want to preserve as constraints, are also modified by the shape approximation.

The existence of prior information such as known landmarks is really valuable in order to solve the aforementioned problems. If these additional priors are introduced and combined with the statistical model they

can be set as constraints in the model. This can be used to obtain initial shapes statistically plausible with the imposed constraints as well as to preserve some shape details during the fitting process.

Researchers have already explored the possibility of introducing additional prior information combined with a statistical model. The most common approach is to introduce a set of known points as prior information. In order to estimate an initial shape from a small set of known points, [Hug et al., 2000] proposed to remove progressively the variation introduced by these points from the PCA model and set them as constraints. However, the statistical model needs to be rebuilt, every time a new point is set as constraint. Following this approach, [Rajamani et al., 2004] extended the method to 3D shapes to reconstruct bones from a small number of control points. Methods such as [Zheng et al., 2006, Rajamani et al., 2007] proposed a least squares error minimization with a regularization term approach. They use a data energy term which minimizes the Euclidean distance between the set of known points and the predicted shape combined with a regularization term which controls the probability of the resulting shape.

Other methods introduced not only known points as prior information but also additional data such as distances between landmarks. [Sierra et al., 2006] presented a method, stated as a constrained minimization problem, to produce viable organ models introducing a set of predictors which consist of variables such as distances measuring the width, length and depth of the organ. This allows to produce new plausible organ instances which meet these shape constraints. [Zhang and Besier, 2016] combined the statistical model with different sparse geometric data for 3D femur reconstruction. The authors use a partial surface point cloud with other measurements as prior information such as the femoral length or other known landmark coordinates. The method minimizes the squared differences between the morphometric measurements or the position of the known landmarks and the corresponding measurements taken on the reconstructed mesh.

[Bernard et al., 2017] have recently addressed the problem of reconstructing the 3D shape with a probabilistic approach based on Gaussian

Mixture Models (GMM), where the known points are interpreted as samples of the GMM.

We propose a least squares minimization with a regularization term approach, similarly to [Zheng et al., 2006, Rajamani et al., 2007]. In our method, instead of only minimizing the least-squares distance of a set of known points [Zheng et al., 2006, Rajamani et al., 2007, Zhang and Besier, 2016], the contribution to the reconstruction error of each point in the shape is controlled by a weights matrix. This allows the possibility of giving more importance to those points whose position is to be preserved (*i.e.*, known landmarks) with respect to the rest of the points in the shape. Moreover, our method requires the computation of the PCA once, differently to the methods [Hug et al., 2000, Rajamani et al., 2004, Sierra et al., 2006] where every time that a point is set as constraint a new PCA needs to be computed and stored removing the variation that the constrained point adds to the statistical model. The WR projection can be used for the task of efficiently initializing a shape from a sparse set of points, providing a statistically plausible shape which preserves the location of these points. Additionally, the weights can be easily adapted to control different regions of points preserving the shape information of more reliable areas during the fitting process.

5.3 Method

In this section, we first present the WR projection is presented. The method is explained in subsection 5.3.1. An extension of the method in order to regularize towards a known shape or a template instead of the mean of the model is explained in subsection 5.3.2. We detail how the WR projection is embedded in the 3DMM fitting method in subsection 5.3.3. Finally we explain in more detail the two considered workflows to obtain the 3D shape reconstruction: 3DMM fitting to scan in subsection 5.3.4 and 3DMM fitting to photos in subsection 5.3.5.

5.3.1 Weighted regularized projection

Our work is based on statistical shape models [Cootes et al., 1995]. A point distribution model (PDM) is used to represent the breast shape. Each 3D shape is formed by a set of n points defined as $\boldsymbol{x} = (X_0, Y_0, Z_0, X_1, Y_1, Z_1, .., X_{n-1}, Y_{n-1}, Z_{n-1})^T$. In order to build the statistical model, a set of training shapes with full correspondence is needed. Once the training shapes are aligned, PCA is applied. This results in a statistical model, where new shape instances can be generated with the following formulation:

$$\boldsymbol{x} = \bar{\boldsymbol{x}} + \boldsymbol{\Phi b} \tag{5.1}$$

where $\bar{\boldsymbol{x}}$ is the mean shape, $\boldsymbol{\Phi}$ is the matrix containing the eigenvectors of the covariance matrix ordered by largest eigenvalue which represents the main modes of variation. The first t main modes can be retained in order to reduce the dimensionality of the problem. The vector containing the eigenvalues is defined as $\boldsymbol{\tau}$. Each eigenvector is weighted by the vector $\boldsymbol{b}$ which represents the shape $\boldsymbol{x}$ in the shape parameter space.

When the 3DMM is fitted to a target shape, the plausibility of the obtained shape needs to be ensured. Under the assumption that given a shape distribution, the subspace of allowed shapes is confined within a hyperellipsoid lying in the shape space which is centered in the mean and has the eigenvectors with largest values as axes, several methods can be applied to approximate a shape to this subspace to ensure its plausibility. Figure 5.2 illustrates the different projection strategies detailed below to approximate a shape into the space of allowed shapes. A simple approach is to approximate to the hyperrectangle space in which the hyperellipsoid is confined [Cootes et al., 1995]. This is done by applying limits to each shape component of the vector b:

$$-D_{max}\sqrt{\boldsymbol{\tau}_i} \leq \boldsymbol{b}_i \leq D_{max}\sqrt{\boldsymbol{\tau}_i} \quad (i = 1, ..., t) \tag{5.2}$$

where D_{max} is a threshold usually set to values between 1.0 and 3.0 which provide suitable shapes. This method offers a great simplification of the shape correction step but some unlikely instances can appear if many values of the vector $\boldsymbol{b}$ are approximated with the maximum value of

$D_{max}\sqrt{\boldsymbol{\tau}_i}$. Alternatively, shapes can be scaled to the surface of the hyperellipsoid in such way that the Mahalanobis distance from the mean D_m is less or equal than a threshold D_{max}:

$$D_m^2 = \sum_{k=1}^{t} \frac{\boldsymbol{b}_k^2}{\boldsymbol{\tau}_k} \leq D_{max}^2 \tag{5.3}$$

The shape vector b can be re-scaled to lie in the hyperellipsoid surface by using

$$\boldsymbol{b} \leftarrow \boldsymbol{b}\frac{D_{max}}{D_m} \tag{5.4}$$

which is more accurate than the hyperrectangle approximation, but it is not necessarily the closest point to the original shape in the hyperellipsoid's surface. [Cerrolaza et al., 2011] defined the hyperelliptical approximation . They propose a method to approximate shapes to the nearest point of the hyperellipsoid minimizing the possible distortions that may appear due to the simplifications of the previous methods. In Figure 5.3 we show in more detail the differences between the different projection strategies. It can be appreciated how the hyperrectangle projection 5.3b provides a suitable shape but some small distortions on the shape appear in the nipples region due to the simplicity of the method. The scaling projection 5.3c corrects the original shape but it is too restrictive loosing shape information. The hyperelliptical 5.3d and WR 5.3e projections give as a result similar suitable shapes but the WR preserves the original position of a set landmarks as evidenced in Table 5.1 where we indicate the errors of the set of landmarks marked in green in Figure 5.2 between the original shape and each one of its correspondent projection strategies.

The aforementioned methods modify the obtained shape parameters to make the shape belong to the model distribution. Nevertheless, if a small set of points is known beforehand (*i.e.*, landmarks previously detected or given as an input) their location should not change but this shape approximation modifies the position of those landmarks which is not desired. The advantage of our WR projection is that the plausibility of the obtained shape and the contribution of each point are controlled in one

unique step which makes it possible to preserve these points obtaining suitable shapes.

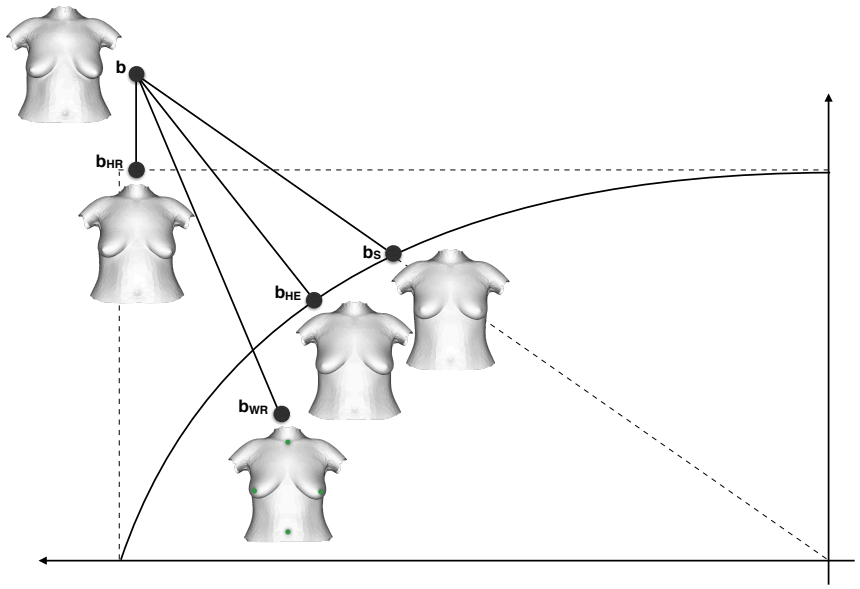
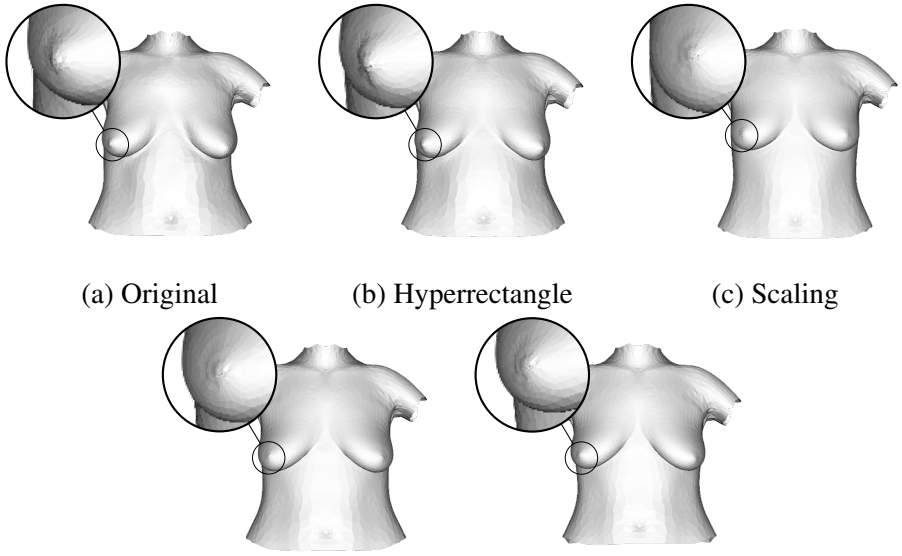


Figure 5.2: The figure shows the different projection strategies to approximate a shape to the allowed subspace of shapes. The shape b represents a shape out of the hyperellipsoid boundary (*i.e.*, not statistically plausible shape). The shape b_{HR} is the approximation to the hyperrectangle using method [Cootes et al., 1995]. The scaling method corresponds to shape b_S. The hyperelliptical approximation [Cerrolaza et al., 2011] is represented by shape b_{HE}. Our weighted regularized projection is represented by b_{WR}, which not necessarily corresponds to the closest point to the initial shape b, but some shape constraints can be fulfilled such as the position of the small set of landmarks marked in green.

We derive our WR projection starting from the work presented by [Zhao et al., 2004] who propose a weighted projection in the shape space. The reconstruction error considering that x^p is the shape space projection



(a) Original (b) Hyperrectangle (c) Scaling

(d) Hyperelliptical (e) WR

Figure 5.3: Projection strategies comparison: the original shape (a) is projected with the different projection strategies.

of x is:

$$ERR = \|\boldsymbol{x} - \boldsymbol{x}^p\|^2 = (\boldsymbol{x} - \boldsymbol{x}^p)^T(\boldsymbol{x} - \boldsymbol{x}^p) \tag{5.5}$$

From this reconstruction error, it is possible to define a point weighted projection performed in the shape space, which introduces another reconstruction error:

$$ERR_W = (\boldsymbol{x} - \boldsymbol{x}^p)^T \boldsymbol{W} (\boldsymbol{x} - \boldsymbol{x}^p) \tag{5.6}$$

where $\boldsymbol{W}$ is a diagonal weights matrix. Our method consists in the optimization of a general energy E which contains two energies, E_D the data term energy and E_R which corresponds to the regularization term energy:

$$E(b) = E_D(\boldsymbol{b}) + \lambda E_R(\boldsymbol{b}) \tag{5.7}$$

where λ is a weight factor to control the contribution of the regularization term. The data term E_D is a weighted projection which minimizes

projection strategies	landmark 1	landmark 2	landmark 3	landmark 4
Hyperrectangle	2.6267	1.8289	1.5031	2.0796
Scaling	4.3240	18.3448	21.9015	23.4623
Hyperelliptical	6.2695	5.4813	7.3347	8.6124
WR	0.0264	0.0310	0.0333	0.01696

Table 5.1: Landmark distances for each projection strategy.

the reconstruction error defined in Equation (5.6) weighting each point contribution:

$$\begin{aligned} E_D(\boldsymbol{b}) &= (\boldsymbol{x} - \boldsymbol{x^p})^T \boldsymbol{W} (\boldsymbol{x} - \boldsymbol{x^p}) = \\ &= (\boldsymbol{x} - \bar{\boldsymbol{x}} - \boldsymbol{\Phi b})^T \boldsymbol{W} (\boldsymbol{x} - \bar{\boldsymbol{x}} - \boldsymbol{\Phi b}) \end{aligned} \tag{5.8}$$

The regularization term corresponds to the Mahalanobis distance to the mean, which will favor the plausibility of the resulting shape:

$$E_R(\boldsymbol{b}) = \boldsymbol{b}^T \boldsymbol{\Lambda}^{-1} \boldsymbol{b} \tag{5.9}$$

where $\boldsymbol{\Lambda}$ corresponds to the matrix of the retained non-zero eigenvalues $\boldsymbol{\Lambda} = diag(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_t)$.

As a result of substituting (5.8) and (5.9) in the general energy Equation (5.7), the energy to minimize is:

$$E(\boldsymbol{b}) = (\boldsymbol{x} - \bar{\boldsymbol{x}} - \boldsymbol{\Phi b})^T \boldsymbol{W} (\boldsymbol{x} - \bar{\boldsymbol{x}} - \boldsymbol{\Phi b}) + \lambda \boldsymbol{b}^T \boldsymbol{\Lambda}^{-1} \boldsymbol{b} \tag{5.10}$$

Then the defined energy $E(\boldsymbol{b})$ is derived to find such shape vector $\boldsymbol{b}$ that

minimizes it:

$$\begin{aligned}\nabla_{\boldsymbol{b}}(E(\boldsymbol{b})) \equiv & \nabla_{\boldsymbol{b}}(-\boldsymbol{x}^T\boldsymbol{W}\boldsymbol{\Phi}\boldsymbol{b} + \bar{\boldsymbol{x}}^T\boldsymbol{W}\boldsymbol{\Phi}\boldsymbol{b} - (\boldsymbol{\Phi}\boldsymbol{b})^T\boldsymbol{W}\boldsymbol{x} + (\boldsymbol{\Phi}\boldsymbol{b})^T\boldsymbol{W}\bar{\boldsymbol{x}} \\ & + (\boldsymbol{\Phi}\boldsymbol{b})^T\boldsymbol{W}\boldsymbol{\Phi}\boldsymbol{b} + \lambda\boldsymbol{b}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{b}) \\ \equiv & \nabla_{\boldsymbol{b}}(-(\boldsymbol{x} - \bar{\boldsymbol{x}})^T\boldsymbol{W}\boldsymbol{\Phi}\boldsymbol{b} - (\boldsymbol{\Phi}\boldsymbol{b})^T\boldsymbol{W}(\boldsymbol{x} - \bar{\boldsymbol{x}}) \\ & + (\boldsymbol{\Phi}\boldsymbol{b})^T\boldsymbol{W}\boldsymbol{\Phi}\boldsymbol{b} + \lambda\boldsymbol{b}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{b}) \\ \equiv & \nabla_{\boldsymbol{b}}(-2(\boldsymbol{x} - \bar{\boldsymbol{x}})^T\boldsymbol{W}\boldsymbol{\Phi}\boldsymbol{b} + (\boldsymbol{\Phi}\boldsymbol{b})^T\boldsymbol{W}\boldsymbol{\Phi}\boldsymbol{b} + \lambda\boldsymbol{b}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{b}) \\ \equiv & \nabla_{\boldsymbol{b}}(-2\boldsymbol{b}^T\boldsymbol{\Phi}^T\boldsymbol{W}^T(\boldsymbol{x} - \bar{\boldsymbol{x}}) + \boldsymbol{b}^T\boldsymbol{\Phi}^T\boldsymbol{W}\boldsymbol{\Phi}\boldsymbol{b} + \lambda\boldsymbol{b}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{b}) \\ \equiv & -2\boldsymbol{\Phi}^T\boldsymbol{W}(\boldsymbol{x} - \bar{\boldsymbol{x}}) + 2\boldsymbol{\Phi}^T\boldsymbol{W}\boldsymbol{\Phi}\boldsymbol{b} + 2\lambda\boldsymbol{\Lambda}^{-1}\boldsymbol{b} = 0 \\ & \boldsymbol{b}(\boldsymbol{\Phi}^T\boldsymbol{W}\boldsymbol{\Phi} + \lambda\boldsymbol{\Lambda}^{-1}) = \boldsymbol{\Phi}^T\boldsymbol{W}(\boldsymbol{x} - \bar{\boldsymbol{x}})\end{aligned} \tag{5.11}$$

The derivation leads to the closed form solution:

$$\boldsymbol{b} = (\boldsymbol{\Phi}^T\boldsymbol{W}\boldsymbol{\Phi} + \lambda\boldsymbol{\Lambda}^{-1})^{-1}\boldsymbol{\Phi}^T\boldsymbol{W}(\boldsymbol{x} - \bar{\boldsymbol{x}}) \tag{5.12}$$

Shape plausibility is controlled by λ which determines the relative importance of the regularization term. Higher values of λ give as a result shapes closer to the mean shape by minimizing the Mahalanobis distance. The contribution of each point in the shape can be weighted with the weights matrix $\boldsymbol{W}$, where higher weights can be applied to those points corresponding to landmarks whose position we want to preserve. This weighted regularized projection allows to obtain plausible shapes and at the same time weigh the contribution of the desired shape points, allowing to give more importance to those points we can rely on.

5.3.2 Prior shape regularization

As described in the previous section, the WR projection controls the shape plausibility regularizing towards the mean shape minimizing the Mahalanobis distance. In some cases, the target shape can be distant from the mean shape, and the regularization term might result too restrictive. We propose to introduce into the formulation the regularization towards an

alternative shape (*i.e.*, a prior shape or a predefined template) instead of using the mean shape of the statistical model. We add to the general energy Equation (5.7), a prior regularization energy term E_P:

$$E(\boldsymbol{b}) = E_D(\boldsymbol{b}) + \lambda E_R(\boldsymbol{b}) + \beta E_P(\boldsymbol{b}) \tag{5.13}$$

where β weights the contribution of the prior regularization energy and $E_P(\boldsymbol{b})$ minimizes the distance error between $\boldsymbol{b}$ and a given prior shape $\boldsymbol{b}_0$:

$$E_P(\boldsymbol{b}) = \|\boldsymbol{b} - \boldsymbol{b_0}\|^2 = (\boldsymbol{b} - \boldsymbol{b}_0)^T(\boldsymbol{b} - \boldsymbol{b}_0) \tag{5.14}$$

where $\boldsymbol{b}_0$ is the vector in the parameter space of the given shape we want to regularize to. The general energy Equation defined in (5.13) is then:

$$E(\boldsymbol{b}) = (\boldsymbol{x} - \bar{\boldsymbol{x}} - \boldsymbol{\Phi b})^T \boldsymbol{W}(\boldsymbol{x} - \bar{\boldsymbol{x}} - \boldsymbol{\Phi b}) + \lambda \boldsymbol{b}^T \boldsymbol{\Lambda}^{-1} \boldsymbol{b} + \beta(\boldsymbol{b} - \boldsymbol{b}_0)^T(\boldsymbol{b} - \boldsymbol{b}_0) \tag{5.15}$$

After deriving we get the following closed form solution:

$$\boldsymbol{b} = (\boldsymbol{\Phi}^T \boldsymbol{W} \boldsymbol{\Phi} + \lambda \boldsymbol{\Lambda}^{-1} + \beta)^{-1}(\boldsymbol{\Phi}^T \boldsymbol{W}(\boldsymbol{x} - \bar{\boldsymbol{x}}) + \beta \boldsymbol{b}_0) \tag{5.16}$$

In this case, higher values of β will favor shapes similar to the given prior shape $\boldsymbol{b}_0$.

5.3.3 3DMM fitting process

During the fitting process we aim to estimate the shape parameters that minimize the difference between the 3DMM shape and the input data. This input can be either a scan or a set of 2D pictures. After pre-processing the input data, a cloud of points $\boldsymbol{d}$ is obtained which are the target data points we want to deform the 3DMM mesh towards (in sections 5.3.4 and 5.3.5 we detail how the data points $\boldsymbol{d}$ are obtained for both types of input: 3D scan and 2D pictures).

The fitting process is based on a non-rigid ICP approach. The deformation applied is a Laplacian deformation with soft constraints [Botsch and Sorkine, 2008], which deforms the mesh towards $\boldsymbol{d}$, while local details are preserved. The 3DMM mesh $\mathcal{M}$ contains n vertices $(\boldsymbol{x}_0, ..., \boldsymbol{x}_{n-1})$.

As defined in [Botsch and Sorkine, 2008], certain vertices can be handled as soft constraints. We consider that the first n' vertices are free while, the $k = n - n'$ last vertices $(\boldsymbol{x}_{n'+1}, ..., \boldsymbol{x}_{n-1})$ are the soft constrained points which we want to deform towards their corresponding points in $\boldsymbol{d}$. The closest point in $\boldsymbol{d}$ to a given point $\boldsymbol{x}_j$ is represented by $cp(\boldsymbol{d}, \boldsymbol{x}_j)$.

We embed our WR projection into the minimization proposed in the work of [Botsch and Sorkine, 2008], since we want to deform the mesh finding the shape parameters subject to being WR projected in order to be able to weight the contribution of each point in the shape to the reconstruction error minimized in Equation (5.12). The energy to minimize is:

$$\begin{aligned} &\min_{b} \quad \sum_{i=0}^{n-1} \parallel \Delta_S(\boldsymbol{x}'_i(\boldsymbol{b})) - \Delta_S(\boldsymbol{x}_i(\boldsymbol{b})) \parallel^2 + \alpha \sum_{j=n'+1}^{n-1} \parallel \boldsymbol{x}'_j(\boldsymbol{b}) - cp(\boldsymbol{d}, \boldsymbol{x}_j) \parallel^2 \\ &\text{subject to} \quad \boldsymbol{b} = (\boldsymbol{\Phi}^T \boldsymbol{W} \boldsymbol{\Phi} + \lambda \boldsymbol{\Lambda}^{-1})^{-1} \boldsymbol{\Phi}^T \boldsymbol{W} (\boldsymbol{x'} - \bar{\boldsymbol{x}}) \end{aligned} \tag{5.17}$$

The first Equation term corresponds to the regularization term, where $\boldsymbol{x'}$ is the deformed mesh and $\Delta_S(\boldsymbol{x}_i)$ is the Laplace-Beltrami operator applied to the point $\boldsymbol{x}_i$. This regularization is applied to all the n mesh vertices. The second term corresponds to the data term in which the distance between the soft constrained mesh vertices and their corresponding data points is minimized. The relative contribution between both terms is controlled by the parameter α. Higher values of α provide a solution closer to the data term and lower values tend towards a Laplacian approximation. Given a linear scalar function $f : \mathcal{M} \rightarrow \mathbb{R}$, the discrete Laplace-Beltrami operator applied at a vertex $\boldsymbol{x}_i$ can be expressed as:

$$\Delta_S f(\boldsymbol{x}_i) = \boldsymbol{e}_i \sum_{\boldsymbol{x}_j \in \mathcal{N}_1(\boldsymbol{x}_i)} \boldsymbol{e}_{ij} (f(\boldsymbol{x}_j) - f(\boldsymbol{x}_i)) \tag{5.18}$$

where $\boldsymbol{x}_j \in \mathcal{N}_1(\boldsymbol{x}_i)$ are the incident one-ring neighbors of $\boldsymbol{x}_i$, $\boldsymbol{e}_i$ are per-vertex normalization weights and $\boldsymbol{e}_{ij} = \boldsymbol{e}_{ji}$ are the edge weights:

$$\boldsymbol{e}_i = \frac{1}{A_i}, \boldsymbol{e}_{ij} = \frac{1}{2}(\cot \gamma_{ij} + \cot \theta_{ij}) \tag{5.19}$$

where γ_{ij} and θ_{ij} are the two angles opposite to an edge $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and A_i is the Voronoi area of vertex $\boldsymbol{x}_i$. Following Equation (5.18) the Laplace-Beltrami operator applied to the whole mesh can be expressed as:

$$\begin{pmatrix} \Delta_S f(\boldsymbol{x}_0) \\ \vdots \\ \Delta_S f(\boldsymbol{x}_{n-1}) \end{pmatrix} = \underbrace{\boldsymbol{M}^{-1}\boldsymbol{L}_s}_{\boldsymbol{L}} \begin{pmatrix} f(\boldsymbol{x}_0) \\ \vdots \\ f(\boldsymbol{x}_{n-1}) \end{pmatrix} \tag{5.20}$$

where $\boldsymbol{M}$ is a diagonal matrix containing normalization weights $\boldsymbol{M}_{ii} = 1/\boldsymbol{e}_i$, and $\boldsymbol{L}_s$ is a symmetric matrix which contains the edge weights $\boldsymbol{e}_{ij}$:

$$(L_s)_{ij} = \begin{cases} -\sum_{\boldsymbol{e}_k \in \mathcal{N}_1(\boldsymbol{x}_i)} \boldsymbol{e}_{ik}, & i = j \\ \boldsymbol{e}_{ij}, & \boldsymbol{x}_j \in \mathcal{N}_1(\boldsymbol{x}_i), \\ 0, & \text{otherwise.} \end{cases} \tag{5.21}$$

The minimum of the energy can be found by solving the system:

$$\begin{aligned} \left[\boldsymbol{L}^T\boldsymbol{L} + \alpha^2 \begin{pmatrix} 0 & 0 \\ 0 & \boldsymbol{I}_k \end{pmatrix}\right] \begin{pmatrix} \boldsymbol{x}_0 \\ \vdots \\ \boldsymbol{x}_{n-1} \end{pmatrix} = \\ \boldsymbol{L}^T \begin{pmatrix} \Delta_S(\boldsymbol{x}_0(\boldsymbol{b})) \\ \vdots \\ \Delta_S(\boldsymbol{x}_{n-1}(\boldsymbol{b})) \end{pmatrix} + \alpha^2 \begin{pmatrix} 0 \\ \vdots \\ 0 \\ cp(\boldsymbol{d}, \boldsymbol{x}_{n'+1}) \\ \vdots \\ cp(\boldsymbol{d}, \boldsymbol{x}_{n-1}) \end{pmatrix} \end{aligned} \tag{5.22}$$

The fitting process is performed in an iterative way, where N deformation iterations are applied. At each iteration, the minimization in Equation (5.17) is solved sequentially. We start by finding correspondences between each vertex $\in (\boldsymbol{x}_{n'+1}, ..., \boldsymbol{x}_n)$ and their closest points in $\boldsymbol{d}$. Then, the mesh is deformed applying the Laplacian deformation method with soft constraints solving Equation (5.22). After that, the WR projection is applied using Equation (5.12) to ensure that the resulting shape parameters meets the imposed shape constraints and is a plausible instance of the statistical model.

5.3.4 Fitting the 3DMM to 3D scans

The objective of this workflow is to reconstruct 3D shape models using depth information. The input data points $\boldsymbol{d}$ are the dense cloud of points obtained from the scan and a small set of landmarks $\boldsymbol{d}_l \in \boldsymbol{d}$ placed by the user in the scan. The scan is a dense cloud of points, however it might present some irregularities such as noise or holes. The idea is to fit a mesh with known topology towards the scan in order to solve these possible issues.

Algorithm 1: FITTING 3DMM TO 3D SCANS METHOD

$Retrieve\ \boldsymbol{b}_0\ from\ dataset\ using\ \boldsymbol{d}_l\ as\ descriptor$;
$Build\ \boldsymbol{W}\ setting\ high\ weights\ to\ \boldsymbol{d}_l$;
$Initialize\ \boldsymbol{b}\ using\ Equation$ (5.16);
for $i \leftarrow 1$ **to** N **do**
    $for\ each\ vertex \in (\boldsymbol{x}_{n'+1}, ..., \boldsymbol{x}_{n-1})\ find\ correspondences\ in\ \boldsymbol{d}$;
    $\boldsymbol{x} \leftarrow Laplacian\ deformation\ using$ (5.22);
    $\boldsymbol{b} \leftarrow WR\ projection(x)\ with$ (5.12);
end

The fitting procedure is explained in Algorithm 1. We aim to obtain a good mesh initialization for the fitting process rather than always initialize the method with the same mesh (*e.g.* predesigned templates or the 3DMM mean shape). A prior shape $\boldsymbol{b}_0$ is retrieved from our set of training shapes used to build the statistical model. We select the most similar shape to the target using the set of points $\boldsymbol{d}_l$ as a descriptor and computing the Eculidean distance. As a result, an initial probable shape approximately fulfilling the imposed shape constraints and regularizing towards the retrieved shape $\boldsymbol{b}_0$ is obtained with Equation (5.16). See Figure 5.5 for some WR initialization examples with a different number of initial points $\boldsymbol{d}_l$. It can be appreciated how with only a sparse set of known points the rest of the shape is predicted by the model inducing the breast type, size and asymmetries.

The initialized mesh is then fitted to the scan following the method mentioned in section 5.3.3. The WR projection is used to ensure that the deformed mesh is statistically plausible and the position of the known landmarks $\boldsymbol{d}_l$ remains almost unchanged.

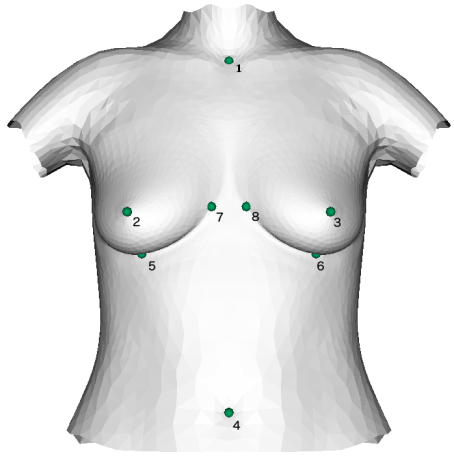


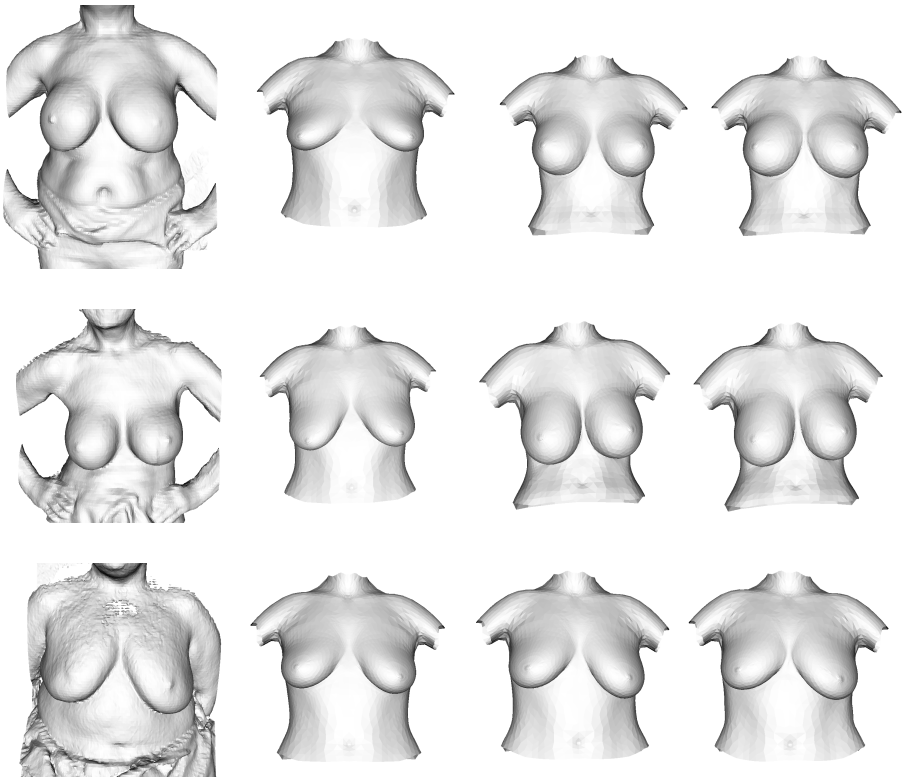
Figure 5.4: Set of initial landmarks which will be used to perform the WR initialization applying high weights on these points. The anatomical positions of these landmarks correspond to the manubrium (1), right nipple (2), left nipple (3), navel (4), right nipple bottom (5), left nipple bottom (6), right inner breast (7), left inner breast (8).

5.3.5 Fitting the 3DMM to 2D pictures

The idea of this workflow is to offer a simple and fast method to obtain the 3D shape of the patient's anatomy when only 2D pictures are available. This is a 3D reconstruction from stereo images problem. The inputs are 3 different pictures, with the only directive of having one frontal picture and one picture of each profile view at $\pm 90^{\circ}$ of the patient. We aim to fit our 3DMM to the 2D pictures and we introduce our WR projection into the shape initialization and shape fitting processes.

Algorithm 2: FITTING 3DMM TO 2D PICTURES METHOD

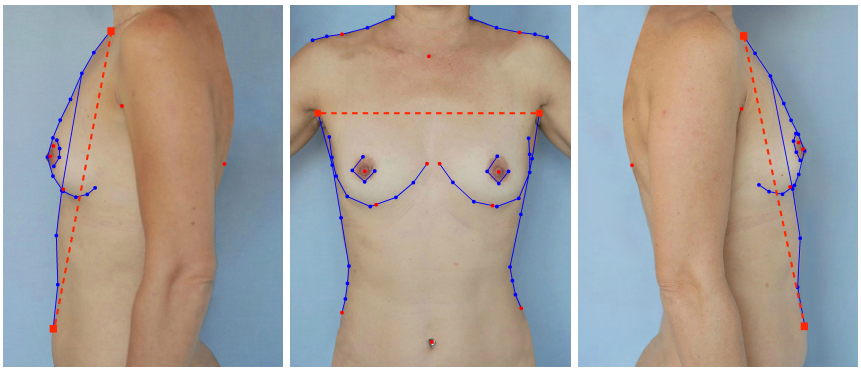
Detect shape $\boldsymbol{d}_{2D}$ *in each view using* [Kazemi and Sullivan, 2014];
Estimation of $\boldsymbol{\kappa}$ *using* $\boldsymbol{d}_{l2D}$ *with resection* [Lepetit et al., 2009];
Compute the $3D$ *positions of* $\boldsymbol{d}_{l2D}$ *to obtain* $\boldsymbol{d}_l$;
Refine $\boldsymbol{\kappa}$ *and* $\boldsymbol{d}_l$ *using bundle adjustment* [Agarwal et al., 2010];
Resample $\boldsymbol{d}_{2D}$ *to match the* $3DMM$ *density*;
Compute the $3D$ *positions of* $\boldsymbol{d}_{2D}$ *to obtain* $\boldsymbol{d}$;
Build $\boldsymbol{W}$ *setting high weights to* $\boldsymbol{d}_l$;
Initialize $\boldsymbol{b}$ *using Equation* (5.12);
for $i \leftarrow 1$ **to** N **do**
    $for\ each\ vertex \in (\boldsymbol{x}_{n'+1}, ..., \boldsymbol{x}_{n-1})\ find\ correspondences\ in\ \boldsymbol{d}$;
    $\boldsymbol{x} \leftarrow Laplacian\ deformation\ using$ (5.22);
    $\boldsymbol{b} \leftarrow WR\ projection(x)\ with$ (5.12);
end



(a) scan (b) 4 constraints (c) 6 constraints (d) 8 constraints

Figure 5.5: Weighted regularized initialization. In this figure we show how with only a set of few known points as constraints (*i.e.*, high weights) the rest of the shape is estimated giving an initial shape closer to the target shape (*i.e.*, the scan). The first column shows the target scans. The second column shows how WR initialization performs with 4 known points (1st to 4th bodymark defined in Figure 5.4). The third column shows WR initialization with 1st to 6th landmarks and last column with the complete set of 8 landmarks.

The fitting procedure is performed as described in Algorithm 2 and ilustrated in Figure 5.7. As a first step, we pre-process the input 2D pictures in order to estimate the set of 3D points $\boldsymbol{d}$ which will be the input



(a) Left view (b) Front view (c) Right view

Figure 5.6: The defined breast shape for each one of the views. In each view the shape is defined as $\boldsymbol{d}_{2D} = \boldsymbol{d}_{b2D} \cup \boldsymbol{d}_{c2D}$ where $\boldsymbol{d}_{l2D}$ are the landmarks plotted in red and $\boldsymbol{d}_{c2D}$ are the contour points plotted in blue. In the frontal view, the number of points in $\boldsymbol{d}_{2D}$ is 52, with 14 landmarks and 38 contour points. In the lateral views the size of $\boldsymbol{d}_{2D}$ is 24 points, with 7 landmarks and 17 contour points. The distances between the two landmarks with square shape marked with a red discontinuous line in each view have been used to normalize the measured errors in the photos database.

data needed to perform the fitting. First of all, the 2D breast shape $\boldsymbol{d}_{2D}$ is detected in each picture using the method from [Kazemi and Sullivan, 2014] (see Figure 5.7 a). This shape is formed by a set of 2D points $\boldsymbol{d}_{2D} = \boldsymbol{d}_{l2D} \cup \boldsymbol{d}_{c2D}$ representing the breast shape, where $\boldsymbol{d}_{l2D}$ are the set of landmarks (*i.e.*, points representing an anatomical point of the body) and $\boldsymbol{d}_{c2D}$ are points defining the shape contour (see Figure 5.6). Using $\boldsymbol{d}_{l2D}$ and the 3DMM mean shape, resection is performed in order to estimate the pose of all cameras $\boldsymbol{\kappa}$ using a state of the art method to solve the PnP problem [Lepetit et al., 2009] (see Figure 5.7 b). Using $\boldsymbol{\kappa}$, the 3D position of each point in $\boldsymbol{d}_{l2D}$ is obtained, which is defined as the set of 3D points $\boldsymbol{d}_l$. The estimates of the camera poses $\boldsymbol{\kappa}$ and $\boldsymbol{d}_l$ are refined in a

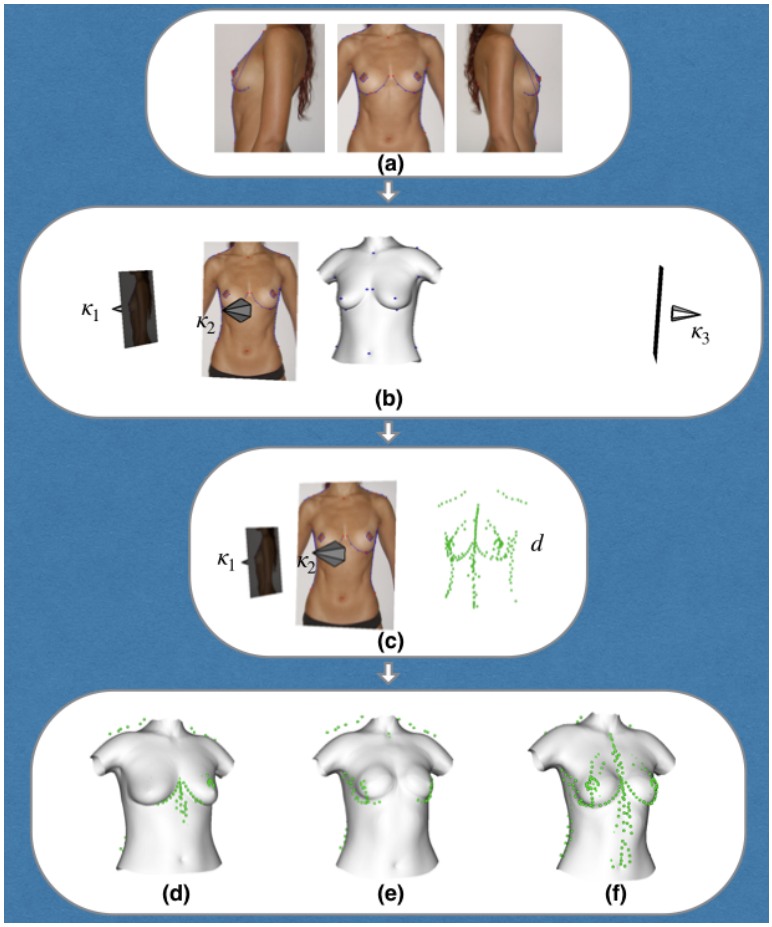


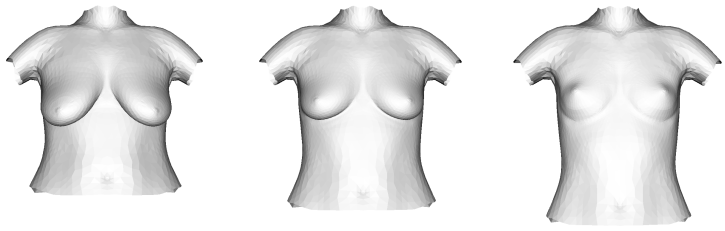
Figure 5.7: 3DMM fitting to 2d pictures. (a) breast shapes are detected in each view picture. (b) Cameras κ are estimated using resection. The method uses the correspondences between the landmarks detected in the 2D pictures (rendered in red) and the 3D landmarks in the mesh (rendered in blue). (c) The 2D contours are projected to 3D resulting in the input data points $\boldsymbol{d}$ rendered in green. (d) Rough initialization using the mean shape. (e) Weighted regularized intialization. (f) 3DMM fitted to the input data poinst $\boldsymbol{d}$ after six iterations.

joint optimization by applying bundle adjustment [Agarwal et al., 2010]. The points in $\boldsymbol{d}_{2D}$ are resampled in order to obtain a point density that matches the one of the 3DMM mesh. For each point in the resampled $\boldsymbol{d}_{2D}$ its 3D position is obtained using $\boldsymbol{\kappa}$ which will result in our input data points $\boldsymbol{d}$ (see Figure 5.7 c). We use our WR projection to obtain an initial 3D shape statistically plausible giving high weights to $\boldsymbol{d}_l$ (see Figure 5.7 e) . This step provides a suitable initialization which gives shapes closer to the target instead of performing a rough initialization always using the same shape (*e.g.* mean shape, predefined template Figure 5.7 d). The initialized mesh is deformed to fit the input data points $\boldsymbol{d}$. The fitting process detailed in section 5.3.3 is used iteratively until convergence. During the fitting higher weights to the points in the 3DMM in correspondence with $\boldsymbol{d}$ are applied in order to preserve the shape of the patient's anatomy extracted from the detection and at the same time obtaining a statistically plausible shape. An example of the fitted shape is shown in Figure 5.7 f.

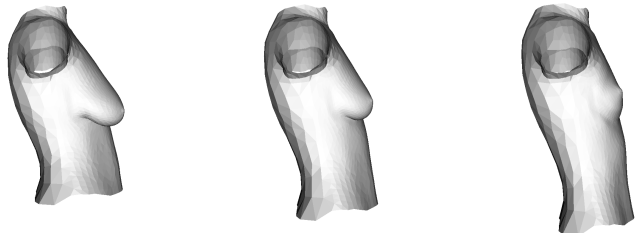
5.4 Experiments

5.4.1 Databases and model generation

In this section several experiments have been carried out in order to evaluate the quality of the 3DMM fitting. In order to assess the two proposed frameworks, two different databases have been used to perform the experiments: scans and photos databases. All the collected data relative to the scans and photos databases has been collected by physicians at their consultations, using Crisalix's software and the structured light Structure Sensor by Occipital in the case of the scans database. Conditions of usage of Crisalix by physicians allow to use the collected data in a completely anonymous way for research and development purposes. There is no personal information about the patient nor the physician the data collected is coming from. The scans database has been used to build the statistical model and test the fitting 3DMM to 3D scans framework (section 5.3.4). The fitting 3DMM to 2D pictures (section 5.3.5) framework has been tested with the photos database. The content of both databases is



(a) mean +2SD (b) mean (c) mean -2SD



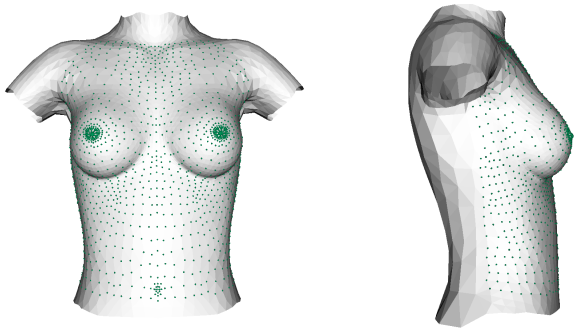
(d) mean +2SD (e) mean (f) mean -2SD

Figure 5.8: Statistical model of the breast anatomy. We show the first mode of variation from a front and lateral point of view. On the left, (a) and (d) show mean +2SD. In the centar (b) and (e) show the mean shape. The mean -2SD can be seen in (c) and (f).

described below:

Scans database and statistical model: The scans database consists of 310 scans obtained by physicians from different patients using Crisalix's software and the structured light scanner Structure Sensor by Occipital. From the point of view of the utility of the application which is the simulation of plastic surgery procedures (*i.e.*, changes in the breast shape and volume) we are interested in the information related with the breast type and size shown in Table 5.2.

To generate the 3DMM we use a reference mesh that represents the breast shape and contains the shape information (i.e., each mesh vertex contains the X, Y, Z coordinates of each shape point), it is formed by 3354 vertices and is symmetric. This mesh is fitted to each scan in the



(a) Front view (b) Lateral view

Figure 5.9: The subset of mesh points lighted in green are used to find point to point correspondences with the scan and compute the distance error.

breast size / type	ptotic	normal	total	percentage
flat	-	16	16	5.16%
small	82	104	186	60.0%
medium	61	22	83	26.78%
big	22	3	25	8.06%
total	165	145	310	

Table 5.2: Scans database population for each breast type and size.

database. The fact of registering a reference mesh to each scan sample in the database provides advantages such as that the parameterization and topology of the different fitted meshes is always the same. Additionally, even if the scan contains holes or missing parts, the fitted mesh will always contain the complete shape with the same number of vertices. Another great advantage is that there are full vertex correspondences between the meshes which allows to easily align the shapes and generate a statistical model with all the shapes. As first step to perform the mesh registration against the scans, 24 landmarks were manually annotated by

experts in the scans; these landmarks which were only used in the 3DMM generation. After the fitting process, the mirrored versions of the registered meshes were generated to increase the variability of the model. To perform the experiments we have used the 310 instances with cross validation using 10 folds. We generated the 10 different statistical models corresponding to each fold using the different combinations of training and testing sets with 90% of instances used for training and leaving the rest for testing. An example of the first mode of variation of the statistical model is shown in Figure 5.8.

The error measures shown in the experiments below are computed by finding correspondences between the fitted meshes and the scans which are used as ground truth. The correspondences are found performing a kd-tree search from each mesh point to the closest scan point. Since the scans are obtained in a frontal perspective from the patient and we do not have data of the patient's back, we consider only the frontal subset of the mesh points (see Figure 5.9) to compute the error measures in the experiments.

Photos database: The photos database consists of 510 samples. The photos are acquired by physicians at their medical consultations. For each sample we have 3 photos: a frontal photo and one photo of each profile at $\pm 90^{\circ}$ of the subject taken with the same camera. However, there is no controlled setup nor calibrated equipment and the pictures of diverse patients are taken with different camera devices and in different medical consultations The breast shapes have been manually labeled by expert assistants for each view as defined in Figure 5.6. The information of the database population by breast size and type is shown in Table 5.3. The database is used to evaluate the 3D reconstruction from photos framework.

5.4.2 3D morphable model initialization

As mentioned above, the initial shape is important for the correct convergence of the fitting process. If this initialization is not good enough, the algorithm can get stuck in local minima. In this experiment, we aim to

breast size / type	ptotic	normal	total	percentage
flat	-	108	108	21.18%
small	58	168	226	44.31%
medium	36	86	122	23.92%
big	16	38	54	10.59%
total	110	400	510	

Table 5.3: Photos database population for each breast size and type.

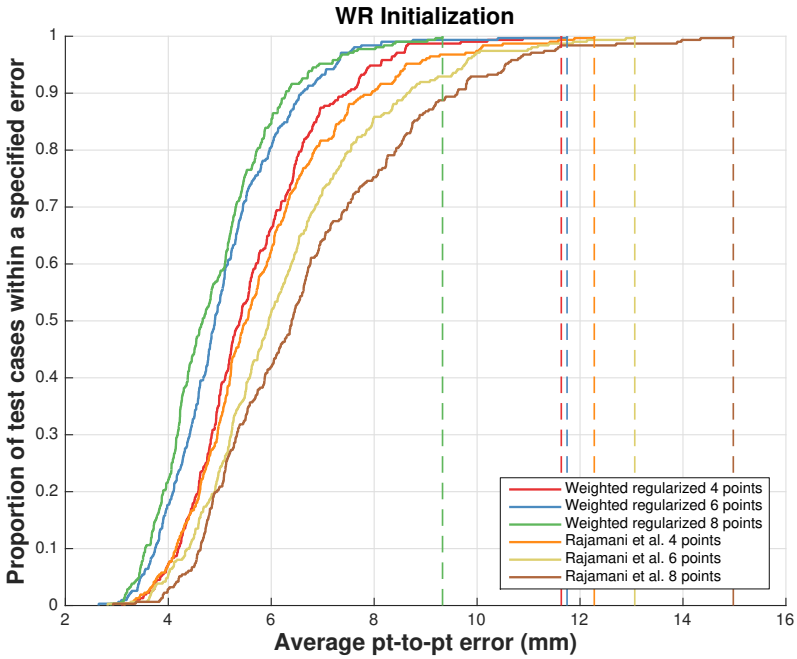


Figure 5.10: 3DMM shape initialization accuracy comparison between [Rajamani et al., 2004] and WR projection with different number of constrained points.

obtain initial meshes that are sufficiently close to the target scan by introducing the additional prior knowledge of a small set of known landmarks

to test the performance of the WR initialization by varying the number of available constraints up to a minimum range.

We consider the set of 8 initial known landmarks detailed in Figure 5.4 and we use them to initialize a mesh with the WR initialization. The known landmarks are set as constraints by applying high weights on them. We have experimentally chosen $\boldsymbol{W}_i = 100$ for those vertex indexes corresponding to the initial points in the weights matrix $\boldsymbol{W}$. For the rest of points in $\boldsymbol{W}$ the weight value is set to 1. We compare our WR initialization with another mesh initialization method proposed by [Rajamani et al., 2004]. From the existing initialization methods, this one is the most similar to ours since it ensures the preservation of a set of known landmarks. It does so by iteratively removing the shape information coded by these landmarks from the statistical model.

We use the scans database to evaluate the initializations using the scans as ground truth. Figure 5.10 shows the mean average error of the initialization using different number of constraints for both methods. As it can be observed, the cummulative error distribution (CED) curve is similar for both methods when initializing with 4 points, although the maximum error using WR is 11.36 mm whereas using [Rajamani et al., 2004] the maximum error is 12.27 mm. However, when the number of initial points is increased using 6 and 8 points to initialize the shape, the WR projection behaves better achieving a lower error. On the other hand, initializing with [Rajamani et al., 2004] and increasing the number of initial landmarks entails a higher error distance of the initialized shape to the target mesh. This is due to the limited variability of the statistical model, given that the method iteratively removes the variability from the statistical model that corresponds to each constrained point. We show in Figure 5.5 some qualitative examples of WR initializations varying the number of initial known points. The results show how suitable initializations can be obtained with the WR initialization and how the method provides verisimilar shapes with just a few known points.

method	mean	median	Std. Dev.	max
weighted regularized	2.36	2.31	0.18	12.67
baseline	3.05	2.89	0.61	15.40

Table 5.4: The table shows statistics of the average distance error of the mesh vertices to the scan for all the reconstructions comparing the two methods: WR projection and baseline. Distances are given in mm.

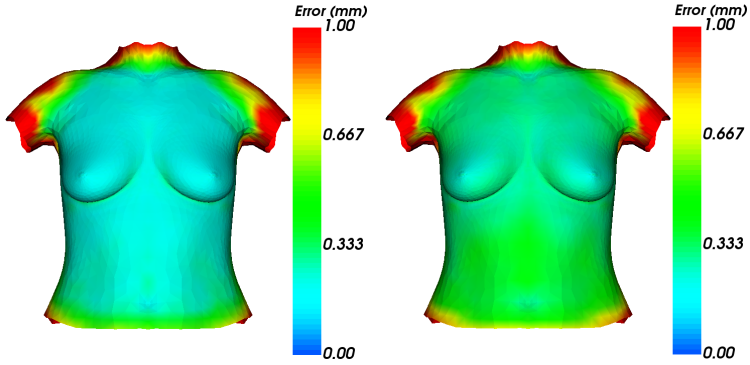
5.4.3 3DMM fitting to scan with WR projection

In this section we evaluate the impact of our WR projection in the process of fitting a 3DMM to a scan in the scans database. We implemented our framework in C++ with parallel CPU execution. The experiments were done in a 3.4GHz Intel Core i7 (16GB RAM) and it took about 16 minutes to execute the 3DMM to all the scans database (310 cases) which gives an average time of 3.15 seconds for each instance. We compare the impact of using our WR projection in the initialization and fitting steps as detailed in section 5.3.4 against using a predefined template as initialization and the statistical model to ensure the shapes plausibility without our WR projection during the registration process, applying hyperelliptical correction [Cerrolaza et al., 2011]. We refer to this last method as baseline. Table 5.4 gives error statistics between the fitted surface meshes and the scans. Due to the impact of using the WR projection in the initialization and fitting steps, the mean average fitting error is 2.36 mm while using the baseline produces 3.05 mm of mean average fitting error. Figure 5.11 shows the mean mesh to scan error color-coded for each vertex on a sample mesh for both methods WR and baseline.

To quantify the benefits of WR projections in the different processing steps, we also tested some partial implementations at half-way between the baseline and WR methods. The Euclidean distance average error comparison between the different variations of the methods is presented in Figure 5.12. As can be observed the baseline is the method with the highest error. When initializing with WR initialization instead of using

a predefined template (WR initialization + baseline) the error is slightly decreased. The advantage is clearly seen when using WR, since the average error decreases and the average maximum error reduced more than 2 mm compared to the baseline. Note the effect of λ in the WR projection. When λ is increased the regularization is stronger which results in more plausible shapes, closer to the mean shape but more distant to the target. In order to see this effect with more detail and analyze the plausibility of the obtained shapes, we show in Figure 5.13 the histogram of the Mahalanobis distances to the mean for all the shapes comparing the previously mentioned methods.

Some fitted examples comparing qualitatively the WR and baseline methods are shown in Figure 5.14. Both methods provide plausible shapes and satisfactory results. Nevertheless the shapes with WR are fitted to the scan more accurately as can be observed in the color coded error distances.



(a) WR (b) baseline

Figure 5.11: Visualization of the mean error distances between the mesh and the scan for WR and baseline methods.

Another advantage of the WR projection is that the weights of each mesh point can be adjusted, giving higher weights to those points that we can rely on, either because they have been previously detected or located

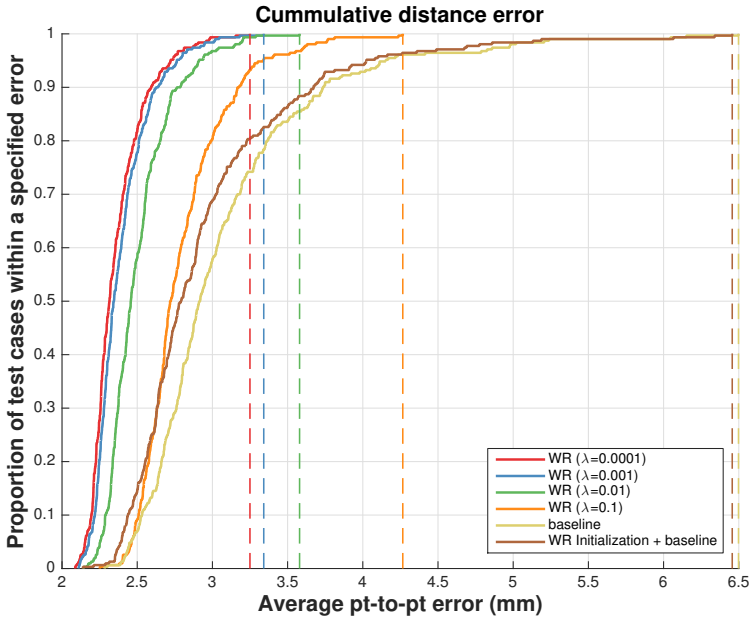


Figure 5.12: Cummulative distance error across all the scan cases. The average distance error is plotted against the test cases population for methods WR with different λ values, the baseline and using WR initialization + baseline.

by the user of the application. Since we previously know the position of 4 landmarks high weights can be applied on them to preserve their location. We measured the error of these landmarks after the fitting process comparing them with the baseline method in Figure 5.15. For this measurement, we applied $\boldsymbol{W}_i = 100$ for those indexes corresponding to the 4 landmarks in the weights matrix $\boldsymbol{W}$. For the rest of points in the shape the weight value was set to 1. It can be seen how using WR projection, after all the fitting process, the position of the known landmarks remains almost unchanged with an average error ≤ 1.01 mm while using the baseline method does not ensure their position with an the average error $\geq$ 2.11 mm for the mentioned landmarks.

However regions such as the inframammary folds in ptotic cases are

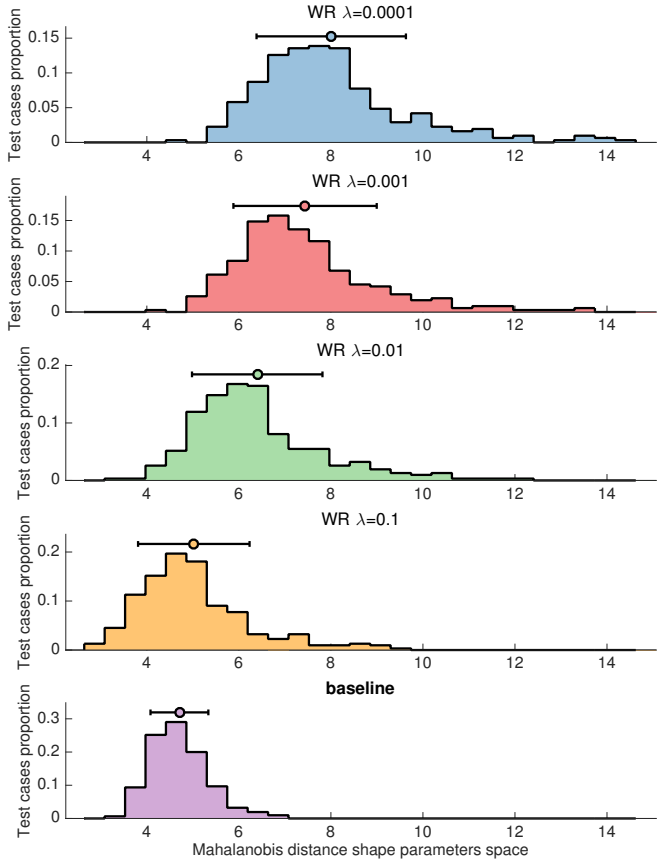
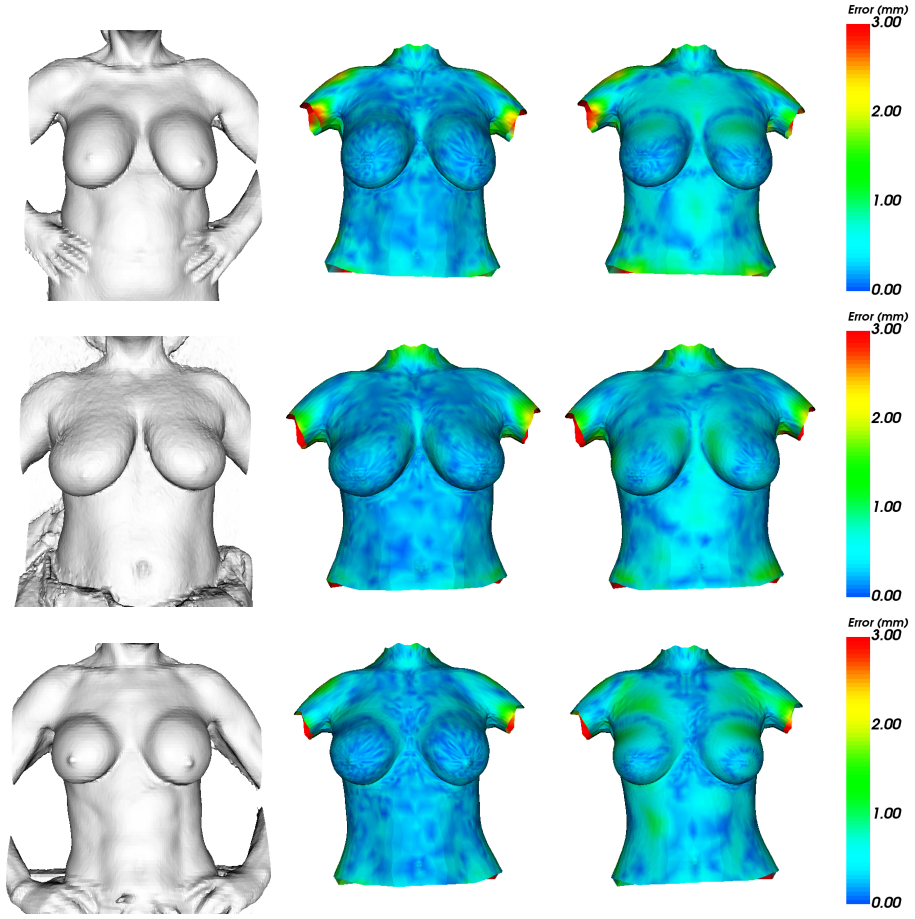


Figure 5.13: Histogram of the Mahalanobis distance to the mean in the shape parameters space. The lines on the top of each plot show the mean and standard deviation of each distribution.

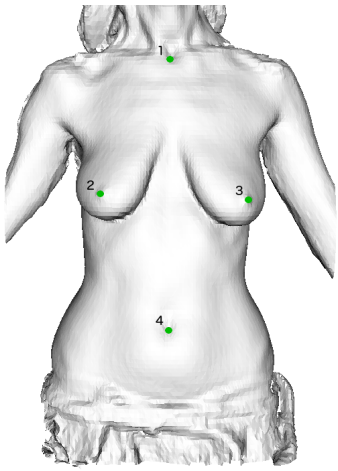
challenging since they remain occluded during the acquisition process, thus the concrete information of these parts of the body are missing. The determination of the inframammary folds using the presented method is based on the statistical information available in the 3DMM. The concrete assessment of the accuracy in the determination of the inframammary folds has not been possible due to the mentioned limitations. For future work this issue could be improved conducting a more in depth study of



(a) scan (b) WR fitting (c) baseline fitting

Figure 5.14: Mesh registration towards a scan. The point to point error distances of each mesh point to the nearest scan point are displayed in colors. Each row shows from left to right: the input scan, the fitted mesh using WR projection, the fitted mesh using baseline method and the legend for the color error distances.

this region, as done in [Henseler et al., 2013] where the patients where told to adopt special poses nearly horizontal in order to eliminate the



	1	2	3	4
WR	0.80	1.01	1.00	0.86
baseline	2.78	2.20	2.11	3.49

Figure 5.15: Landmarks placed in the scan used in the initialization and fitting process on the left. Note their anatomical position that correspond to the manubrium (1), right nipple (2), left nipple (3) and the navel (4). The table shows the average Euclidean error after the fitting, for the 4 landmarks defined on the left, using our weighted regularized method (WR) and the baseline. The error is given in mm.

ptotic effect during the acquisition process.

5.4.4 3DMM fitting to 2D pictures with WR projection

The second component of our framework is a method which allows the fitting of a 3DMM to multi-view 2D photos. We measure how accurately our 3DMM is fitted to the images. To do so, our method is evaluated in the photos database. The 3DMM to photos method takes about 138 minutes to be executed on all the photos database (510 cases) which gives an average of 16.27 seconds per instance. The error measure is the mean contour reprojection distance. In each fitting iteration the mesh contours are projected to the three images (*i.e.*, front, left right views) to compute the distance to the manually labeled contours which are used as ground truth (see Figure 5.16).

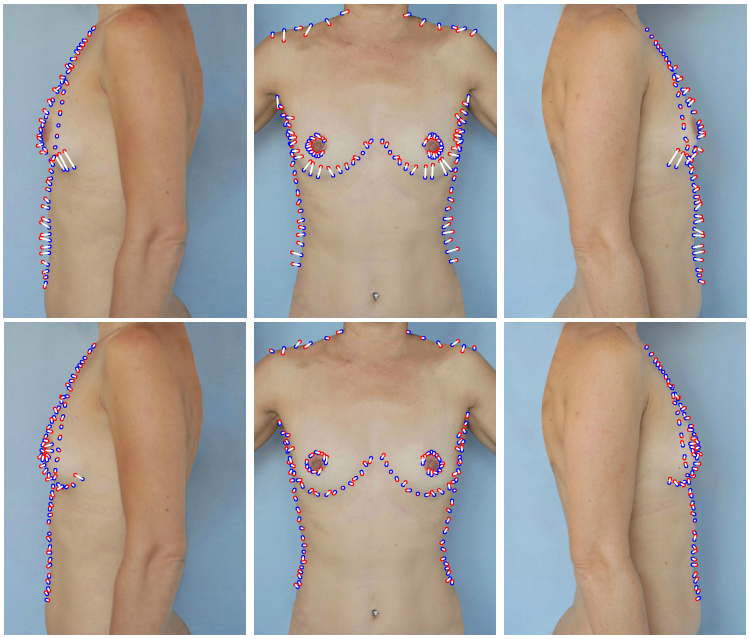


Figure 5.16: Reprojected contours and deformation correspondences. The points in blue are the ground truth contour points. The points in red are the found mesh point correspondences projected in the image. The white lines link each mesh projected point with its correspondent ground truth contour point. The distances between these points are used to compute the contours mean distance error. In the first row, an example of the mesh projected correspondences is shown for the first iteration of the fitting process. The second row shows the correspondences in the last iteration.

We compare the impact of using our WR projection in the initialization and fitting steps as detailed in section 5.3.5 against the baseline method defined in the previous section 5.4.3. In addition, in order to measure the impact of the WR initialization, we also tested some partial implementations at half-way between the baseline and WR methods. The reprojection contour errors comparing the aforementioned methods are shown in Figure 5.17. The results show how, in the initial fitting itera-

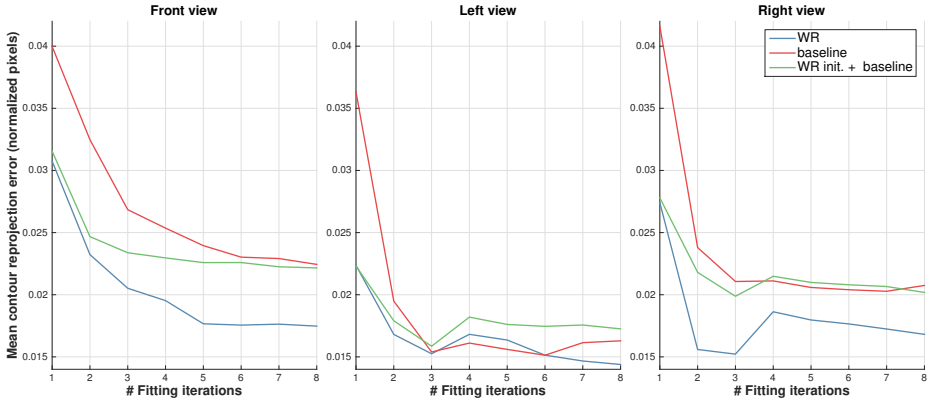


Figure 5.17: Mean contour reprojection error (*i.e.*, error between the ground truth contour points and their correspondent mesh points projected into the image) for each iteration in the fitting method. The error is shown for the three different views (*i.e.*, front, left, right). Given that the images are acquired at different resolutions the pixel error measurements have been normalized using the previously defined distances in Figure 5.6.

tions when using WR intialization the error is considerably lower in the 3 views, than using the baseline which initializes always with the same predefined shape. After several fitting iterations the baseline methods give a similar final error. However, using WR projection during the fitting yields a lower error in the 3 views and therefore a better fitting result.

We show some qualitative results of the 3DMM fitting to the pictures in Figure 5.18. It can be seen how the 3DMM is fitted to the different kind of breasts shown in the examples with satisfactory results. However, given that the camera pose estimation is performed in a joint three-view error minimization, some overlapping errors can be appreciated in certain projected views. The errors appear especially in areas without contours information such as the patient's back, the upper arms or occluded elements, for example the left breast in the projection of the right view. To improve the fitting accuracy we provide the aforementioned 3DMM fitting to 3D scans method.

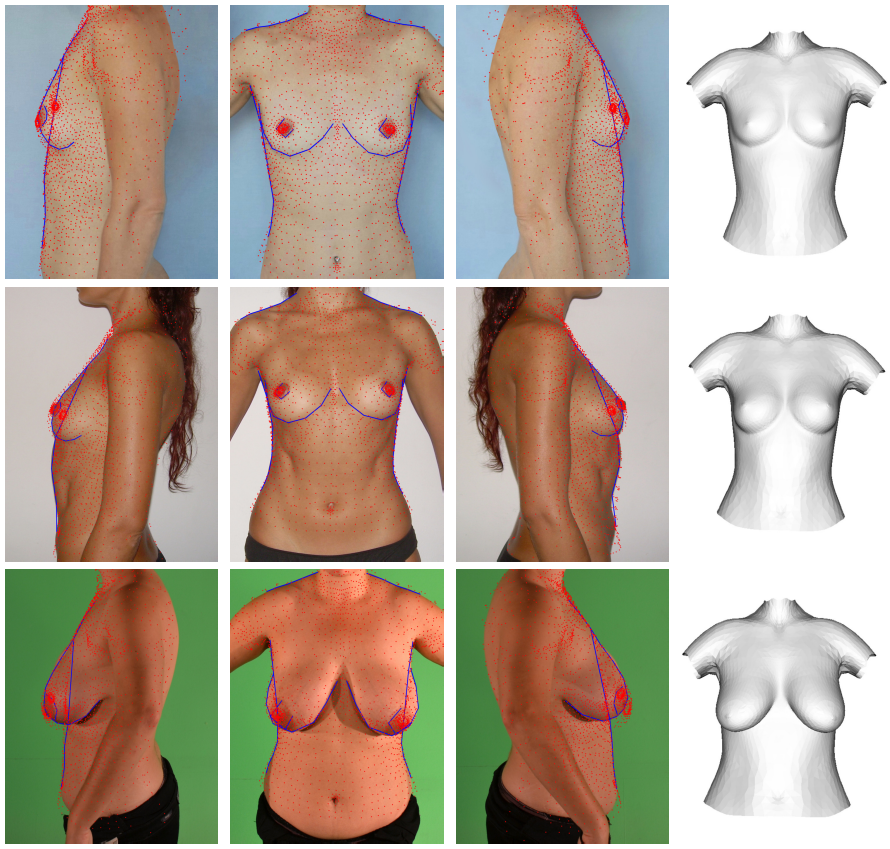


Figure 5.18: 3DMM fitting to 2d photos. The first three columns show the patient pictures. The lines in blue show the ground truth contours and the red points are the obtained fitted mesh points projected onto the image. The last column shows the obtained fitted mesh.

5.5 Conclusions

We have presented a WR projection in the parameter space which allows to impose additional constraints into a statistical model. We have applied this method in a framework capable to fit a 3DMM representing the breast shape to two possible inputs: scans and 2D photos. The presented

framework enables the reconstruction relying on the data of the patient's anatomy without the use of expensive hardware. The output reconstructions are suitable to be used with visualization and simulation purposes during the surgery planning process.

The formulation of the method is flexible and allows to adjust if most of the shape prediction is performed from the statistical model (*i.e.*, regularization term) or from the additional prior information (*i.e.*, data term) and weight the contribution of each point in the shape to the reconstruction error. In this way, statistically plausible shapes with additional constraints introduced into the statistical model are obtained in a unique formulation. The proposed WR projection has been proved to be useful to efficiently initialize shapes from a sparse set of known points and has been compared to the most similar competing method for 3D shape initialization [Rajamani et al., 2004] obtaining better performance. As shown in the results, these initializations predict important shape information such as the breast size or breast type with the only information of a sparse set of known points supported by the statistical model, providing a good starting point for the 3DMM fitting process. In addition, the WR projection has been embedded into the fitting process in order to preserve the shapes plausibility and maintain the position of the constrained 3D points. The results show how the performance of the fitting process is also improved with the usage of the WR projection achieving 3D shapes with lower reconstruction error in both fitting frameworks: scans and 2D photos.

Chapter 6 have been removed from the thesis due to confidentiality issues.

BIBLIOGRAPHY

[Agarwal et al., 2010] Agarwal, S., Snavely, N., Seitz, S. M., and Szeliski, R. (2010). Bundle adjustment in the large. In *European Conference on Computer Vision (ECCV)*, pages 29–42. Springer.

[Aldrian and Smith, 2010] Aldrian, O. and Smith, W. (2010). A linear approach of 3d face shape and texture recovery using a 3d morphable model. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 75–1.

[Aldrian and Smith, 2013] Aldrian, O. and Smith, W. A. (2013). Inverse rendering of faces with a 3d morphable model. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 35(5):1080–1093.

[Alexander et al., 2009] Alexander, O., Rogers, M., Lambeth, W., Chiang, M., and Debevec, P. (2009). Creating a photoreal digital actor: The digital emily project. In *Conference for Visual Media Production (CVMP)*, pages 176–187. IEEE.

[Amberg et al., 2008] Amberg, B., Knothe, R., and Vetter, T. (2008). Expression invariant 3d face recognition with a morphable model. In

International Conference on Automatic Face and Gesture Recognition (FG), pages 1–6. IEEE.

[Angelopoulou and Petrou, 2014] Angelopoulou, M. E. and Petrou, M. (2014). Uncalibrated flatfielding and illumination vector estimationfor photometric stereo face reconstruction. *Machine vision and applications*, 25(5):1317–1332.

[Asthana et al., 2011] Asthana, A., Marks, T. K., Jones, M. J., Tieu, K. H., and Rohith, M. (2011). Fully automatic pose-invariant face recognition via 3d pose normalization. In *International Conference on Computer Vision (ICCV)*, pages 937–944. IEEE.

[Atick et al., 1996] Atick, J. J., Griffin, P. A., and Redlich, A. N. (1996). Statistical approach to shape from shading: Reconstruction of three-dimensional face surfaces from single two-dimensional images. *Neural computation*, 8(6):1321–1340.

[Beeler et al., 2010] Beeler, T., Bickel, B., Beardsley, P., Sumner, B., and Gross, M. (2010). High-quality single-shot capture of facial geometry. In *Transactions on Graphics (TOG)*, volume 29, page 40. ACM.

[Beeler et al., 2011] Beeler, T., Hahn, F., Bradley, D., Bickel, B., Beardsley, P., Gotsman, C., Sumner, R. W., and Gross, M. (2011). High-quality passive facial performance capture using anchor frames. In *Transactions on Graphics (TOG)*, volume 30, page 75. ACM.

[Bernard et al., 2017] Bernard, F., Salamanca, L., Thunberg, J., Tack, A., Jentsch, D., Lamecker, H., Zachow, S., Hertel, F., Goncalves, J., and Gemmar, P. (2017). Shape-aware surface reconstruction from sparse 3D point-clouds. *Medical Image Analysis*.

[Besl and McKay, 1992] Besl, P. J. and McKay, N. D. (1992). Method for registration of 3-D shapes. In *Robotics-DL tentative*, pages 586–606. International Society for Optics and Photonics.

[Blanc et al., 2009] Blanc, R., Syrkina, E., and Székely, G. (2009). Estimating the confidence of statistical model based shape prediction. In *Information Processing in Medical Imaging (IPMI)*, pages 602–613. Springer.

[Blanz and Vetter, 1999] Blanz, V. and Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In *Proceedings of the Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 187–194. ACM Press/Addison-Wesley Publishing Co.

[Blanz and Vetter, 2003] Blanz, V. and Vetter, T. (2003). Face recognition based on fitting a 3D morphable model. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 25(9):1063–1074.

[Botsch and Sorkine, 2008] Botsch, M. and Sorkine, O. (2008). On linear variational surface deformation methods. *Transactions on Visualization and Computer Graphics (TVCG)*, 14(1):213–230.

[Bradley et al., 2010] Bradley, D., Heidrich, W., Popa, T., and Sheffer, A. (2010). High resolution passive facial performance capture. In *Transactions on Graphics (TOG)*, volume 29, page 41. ACM.

[Brunton et al., 2014] Brunton, A., Salazar, A., Bolkart, T., and Wuhrer, S. (2014). Review of statistical shape spaces for 3D data with comparative analysis for human faces. *Computer Vision and Image Understanding*, 128:1–17.

[Cao et al., 2014] Cao, C., Weng, Y., Zhou, S., Tong, Y., and Zhou, K. (2014). Facewarehouse: A 3d facial expression database for visual computing. *Transactions on Visualization and Computer Graphics (TVCG)*, 20(3):413–425.

[Cerrolaza et al., 2011] Cerrolaza, J. J., Villanueva, A., and Cabeza, R. (2011). Shape constraint strategies: Novel approaches and comparative robustness. In *British Machine Vision Conference (BMVC)*, pages 1–11.

[Chang et al., 2015] Chang, J. B., Small, K. H., Choi, M., and Karp, N. S. (2015). Three-dimensional surface imaging in plastic surgery: foundation, practical applications, and beyond. *Plastic and reconstructive surgery*, 135(5):1295–1304.

[Chen et al., 2010] Chen, D., Chittajallu, D., Passalis, G., and Kakadiaris, I. (2010). Computational tools for quantitative breast morphometry based on 3D scans. *Annals of Biomedical Engineering*, 38(5):1703–1718.

[Cootes and Taylor, 2001] Cootes, T. F. and Taylor, C. J. (2001). Constrained active appearance models. In *International Conference on Computer Vision (ICCV)*, volume 1, pages 748–754. IEEE.

[Cootes et al., 1995] Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38–59.

[Cordier et al., 2007] Cordier, F., Hong, K., and Seo, H. (2007). A breast modeler based on analysis of breast scans. *Computer Animation and Virtual Worlds*, 18(2):141–151.

[de Bruijne et al., 2007] de Bruijne, M., Lund, M. T., Tankó, L. B., Pettersen, P. C., and Nielsen, M. (2007). Quantitative vertebral morphometry using neighbor-conditional shape models. *Medical Image Analysis*, 11(5):503–512.

[de Heras Ciechomski et al., 2012] de Heras Ciechomski, P., Constantinescu, M., Garcia, J., Olariu, R., Dindoyal, I., Le Huu, S., and Reyes, M. (2012). Development and implementation of a web-enabled 3D consultation tool for breast augmentation surgery based on 3D-image reconstruction of 2D pictures. *Journal of Medical Internet Research*, 14(1):e21.

[Dou et al., 2017] Dou, P., Shah, S. K., and Kakadiaris, I. A. (2017). End-to-end 3d face reconstruction with deep neural networks. In *Con-*

ference on Computer Vision and Pattern Recognition (CVPR), pages 21–26.

[Dou et al., 2014] Dou, P., Wu, Y., Shah, S. K., and Kakadiaris, I. A. (2014). Robust 3d face shape reconstruction from single images via two-fold coupled structure learning. In *British Machine Vision Conference (BMVC)*, pages 1–13.

[Dovgard and Basri, 2004] Dovgard, R. and Basri, R. (2004). Statistical symmetric shape from shading for 3d structure recovery of faces. In *European Conference on Computer Vision (ECCV)*, pages 99–113. Springer.

[Esteban et al., 2008] Esteban, C. H., Vogiatzis, G., and Cipolla, R. (2008). Multiview photometric stereo. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 30(3):548–554.

[Fyffe et al., 2014] Fyffe, G., Jones, A., Alexander, O., Ichikari, R., and Debevec, P. (2014). Driving high-resolution facial scans with video performance capture. *Transactions on Graphics (TOG)*, 34(1):8.

[Gallo et al., 2010] Gallo, G., Guarnera, G. C., and Catanuto, G. (2010). Human breast shape analysis using pca. In *Biosignals*, pages 163–167. Citeseer.

[Gallo et al., 2009] Gallo, G., Guarnera, G. C., Milanese, F., Modica, D., Catanuto, G., and Pane, F. (2009). Parametric representation of human breast shapes. In *International Workshop on Medical Measurements and Applications (MEMEA)*, pages 26–30. IEEE.

[Geng, 2011] Geng, J. (2011). Structured-light 3d surface imaging: a tutorial. *Advances in Optics and Photonics*, 3(2):128–160.

[Georgii et al., 2014] Georgii, J., Eder, M., Burger, K., Klotz, S., Ferstl, F., Kovacs, L., and Westermann, R. (2014). A computational tool for preoperative breast augmentation planning in aesthetic plastic surgery. *Journal of Biomedical and Health Informatics*, 18(3):907–919.

[Gower, 1975] Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40(1):33–51.

[Häne et al., 2017] Häne, C., Tulsiani, S., and Malik, J. (2017). Hierarchical surface prediction for 3d object reconstruction. In *International Conference on 3D Vision (3DV)*, pages 412–420. IEEE.

[Hansen et al., 2010] Hansen, M. F., Atkinson, G. A., Smith, L. N., and Smith, M. L. (2010). 3d face reconstructions from photometric stereo using near infrared and visible light. *Computer Vision and Image Understanding*, 114(8):942–951.

[Henseler et al., 2013] Henseler, H., Ju, X., Ayoub, A., and Ray, A. K. (2013). The importance of the pose in three-dimensional imaging of the ptotic breast. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, 66(11):1551–1556.

[Henseler et al., 2011] Henseler, H., Khambay, B. S., Bowman, A., Smith, J., Siebert, J. P., Oehler, S., Ju, X., Ayoub, A., and Ray, A. K. (2011). Investigation into accuracy and reproducibility of a 3D breast imaging system using multiple stereo cameras. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, 64(5):577–582.

[Hoeffelin et al., 2014] Hoeffelin, H., Jacquemin, D., Defaweux, V., and Nizet, J. L. (2014). A methodological evaluation of volumetric measurement techniques including three-dimensional imaging in breast surgery. *BioMed Research International*, 2014.

[Horn, 1975] Horn, B. K. (1975). Obtaining shape from shading information. *The psychology of computer vision*, pages 115–155.

[Horn and Brooks, 1989] Horn, B. K. and Brooks, M. J. (1989). *Shape from shading*. MIT press.

[Hu et al., 2017] Hu, G., Yan, F., Kittler, J., Christmas, W., Chan, C. H., Feng, Z., and Huber, P. (2017). Efficient 3d morphable face model fitting. *Pattern Recognition*, 67:366–379.

[Huber et al., 2015] Huber, P., Feng, Z.-H., Christmas, W., Kittler, J., and Ratsch, M. (2015). Fitting 3D morphable face models using local features. In *International Conference on Image Processing (ICIP)*, pages 1195–1199. IEEE.

[Huber et al., 2016] Huber, P., Hu, G., Tena, R., Mortazavian, P., Koppen, W. P., Christmas, W., Rätsch, M., and Kittler, J. (2016). A multiresolution 3D morphable face model and fitting framework. In *International Conference on Computer Vision Theory and Applications (VISAPP)*, pages 1–8.

[Hug et al., 2000] Hug, J., Brechbühler, C., and Székely, G. (2000). Model-based initialisation for segmentation. In *European Conference on Computer Vision (ECCV)*, pages 290–306. Springer.

[Jackson et al., 2017] Jackson, A. S., Bulat, A., Argyriou, V., and Tzimiropoulos, G. (2017). Large pose 3d face reconstruction from a single image via direct volumetric cnn regression. In *International Conference on Computer Vision (ICCV)*, pages 1031–1039. IEEE.

[Jeni et al., 2015] Jeni, L. A., Cohn, J. F., and Kanade, T. (2015). Dense 3d face alignment from 2d videos in real-time. In *International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, volume 1, pages 1–8. IEEE.

[Jiang et al., 2018] Jiang, L., Zhang, J., Deng, B., Li, H., and Liu, L. (2018). 3d face reconstruction with geometry details from a single image. *Transactions on Image Processing*, pages 1–1.

[Jolliffe, 1986] Jolliffe, I. T. (1986). *Principal component analysis and factor analysis*. Springer.

[Jourabloo and Liu, 2016] Jourabloo, A. and Liu, X. (2016). Large-pose face alignment via CNN-based dense 3D model fitting. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4188–4196. IEEE.

[Jourabloo and Liu, 2017] Jourabloo, A. and Liu, X. (2017). Pose-invariant face alignment via cnn-based dense 3d model fitting. *International Journal of Computer Vision*, pages 1–17.

[Kagiyama et al., 2016] Kagiyama, Y., Otomaru, I., Takao, M., Sugano, N., Nakamoto, M., Yokota, F., Tomiyama, N., Tada, Y., and Sato, Y. (2016). Ct-based automated planning of acetabular cup for total hip arthroplasty (tha) based on hybrid use of two statistical atlases. *International journal of computer assisted radiology and surgery*, 11(12):2253–2271.

[Kazemi and Sullivan, 2014] Kazemi, V. and Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1867–1874. IEEE.

[Kemelmacher and Basri, 2006] Kemelmacher, I. and Basri, R. (2006). Molding face shapes by example. In *European Conference on Computer Vision (ECCV)*, pages 277–288. Springer.

[Kemelmacher-Shlizerman and Basri, 2011] Kemelmacher-Shlizerman, I. and Basri, R. (2011). 3d face reconstruction from a single image using a single reference face shape. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 33(2):394–405.

[Kemelmacher-Shlizerman and Seitz, 2011] Kemelmacher-Shlizerman, I. and Seitz, S. M. (2011). Face reconstruction in the wild. In *International Conference on Computer Vision (ICCV)*, pages 1746–1753. IEEE.

[Kim et al., 2017] Kim, Y., Kim, H., and Kim, Y. O. (2017). Virtual reality and augmented reality in plastic surgery: a review. *Archives of plastic surgery*, 44(3):179.

[Kim et al., 2008] Kim, Y., Lee, K., and Kim, W. (2008). 3D virtual simulator for breast plastic surgery. *Computer Animation and Virtual Worlds*, 19(3-4):515–526.

[Kittler et al., 2016] Kittler, J., Huber, P., Feng, Z.-H., Hu, G., and Christmas, W. (2016). 3D morphable face models and their applications. In *International Conference on Articulated Motion and Deformable Objects (AMDO)*, pages 185–206. Springer.

[Kovacs et al., 2006] Kovacs, L., Yassouridis, A., Zimmermann, A., Brockmann, G., Wöhnl, A., Blaschke, M., Eder, M., Schwenzer-Zimmerer, K., Rosenberg, R., Papadopulos, N. A., et al. (2006). Optimization of 3-dimensional imaging of the breast region with 3-dimensional laser scanners. *Annals of Plastic Surgery*, 56(3):229–236.

[Lekakis et al., 2016] Lekakis, G., Claes, P., Hamilton III, G. S., and Hellings, P. (2016). Three-dimensional surface imaging and the continuous evolution of preoperative and postoperative assessment in rhinoplasty. *Facial Plastic Surgery*, 32(01):088–094.

[Lepetit et al., 2009] Lepetit, V., Moreno-Noguer, F., and Fua, P. (2009). Epnp: An accurate o (n) solution to the pnp problem. *International Journal of Computer Vision*, 81(2):155–166.

[Liu et al., 2015] Liu, F., Zeng, D., Li, J., and Zhao, Q. (2015). Cascaded regressor based 3d face reconstruction from a single arbitrary view image. arxiv preprint. *arXiv preprint arXiv:1509.06161*, 2.

[Liu et al., 2016] Liu, F., Zeng, D., Zhao, Q., and Liu, X. (2016). Joint face alignment and 3d face reconstruction. In *European Conference on Computer Vision (ECCV)*, pages 545–560. Springer.

[Liu et al., 2001] Liu, Z., Zhang, Z., Jacobs, C., and Cohen, M. (2001). Rapid modeling of animated faces from video. *The Journal of Visualization and Computer Animation*, 12(4):227–240.

[Milborrow et al., 2010] Milborrow, S., Morkel, J., and Nicolls, F. (2010). The MUCT Landmarked Face Database. *Pattern Recognition Association of South Africa (PRASA)*. urlhttp://www.milbo.org/muct.

[O'Connell et al., 2015] O'Connell, R. L., Stevens, R. J., Harris, P. A., and Rusby, J. E. (2015). Review of three-dimensional (3D) surface imaging for oncoplastic, reconstructive and aesthetic breast surgery. *The Breast*, 24(4):331–342.

[Otomaru et al., 2012] Otomaru, I., Nakamoto, M., Kagiyama, Y., Takao, M., Sugano, N., Tomiyama, N., Tada, Y., and Sato, Y. (2012). Automated preoperative planning of femoral stem in total hip arthroplasty from 3d ct data: Atlas-based approach and comparative study. *Medical image analysis*, 16(2):415–426.

[Paysan et al., 2009] Paysan, P., Knothe, R., Amberg, B., Romdhani, S., and Vetter, T. (2009). A 3d face model for pose and illumination invariant face recognition. In *International Conference on Advanced video and signal based surveillance (AVSS)*, pages 296–301. IEEE.

[Pighin et al., 2006] Pighin, F., Hecker, J., Lischinski, D., Szeliski, R., and Salesin, D. H. (2006). Synthesizing realistic facial expressions from photographs. In *Special Interest Group on Computer Graphics and Interactive Techniques (SIGGRAPH) Courses*, page 19. ACM.

[Rajamani et al., 2004] Rajamani, K. T., Hug, J., Nolte, L. P., and Styner, M. (2004). Bone morphing with statistical shape models for enhanced visualization. In *Medical Imaging*, pages 122–130. International Society for Optics and Photonics.

[Rajamani et al., 2007] Rajamani, K. T., Styner, M. A., Talib, H., Zheng, G., Nolte, L. P., and González Ballester, M. A. (2007). Statistical deformable bone models for robust 3D surface extrapolation from sparse data. *Medical Image Analysis*, 11(2):99–109.

[Richardson et al., 2016] Richardson, E., Sela, M., and Kimmel, R. (2016). 3d face reconstruction by learning from synthetic data. In *International Conference on 3D Vision (3DV)*, pages 460–469. IEEE.

[Richardson et al., 2017] Richardson, E., Sela, M., Or-El, R., and Kimmel, R. (2017). Learning detailed face reconstruction from a single

image. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5553–5562. IEEE.

[Romdhani and Vetter, 2005] Romdhani, S. and Vetter, T. (2005). Estimating 3d shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 986–993. IEEE.

[Roth et al., 2015] Roth, J., Tong, Y., and Liu, X. (2015). Unconstrained 3d face reconstruction. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2606–2615. IEEE.

[Roth et al., 2016] Roth, J., Tong, Y., and Liu, X. (2016). Adaptive 3d face reconstruction from unconstrained photo collections. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4197–4206. IEEE.

[Ruiz et al., 2018] Ruiz, G., Ramon, E., García, J., Sukno, F. M., and Ballester, M. A. G. (2018). Weighted regularized statistical shape space projection for breast 3d model reconstruction. *Medical image analysis*, 47:164–179.

[Ruiz et al., 2016] Ruiz, G., Ramon, E., González Ballester, M. A., Sukno, F. M., et al. (2016). Weighted regularized ASM for face alignment. In *International Conference on Image Processing (ICIP)*, pages 2906–2910. IEEE.

[Salvi et al., 2010] Salvi, J., Fernandez, S., Pribanic, T., and Llado, X. (2010). A state of the art in structured light patterns for surface profilometry. *Pattern recognition*, 43(8):2666–2680.

[Salvi et al., 2004] Salvi, J., Pages, J., and Batlle, J. (2004). Pattern codification strategies in structured light systems. *Pattern recognition*, 37(4):827–849.

[Seitz et al., 2006] Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 519–528. IEEE.

[Sela et al., 2017] Sela, M., Richardson, E., and Kimmel, R. (2017). Unrestricted facial geometry reconstruction using image-to-image translation. In *International Conference on Computer Vision (ICCV)*, pages 1585–1594. IEEE.

[Shimshoni et al., 2000] Shimshoni, I., Moses, Y., and Lindenbaum, M. (2000). Shape reconstruction of 3d bilaterally symmetric surfaces. *International Journal of Computer Vision*, 39(2):97–110.

[Sierra et al., 2006] Sierra, R., Zsemlye, G., Székely, G., and Bajka, M. (2006). Generation of variable anatomical models for surgical training simulators. *Medical Image Analysis*, 10(2):275–285.

[Smith and Zhang, 2014] Smith, B. M. and Zhang, L. (2014). Collaborative facial landmark localization for transferring annotations across datasets. In *European Conference on Computer Vision (ECCV)*, pages 78–93. Springer.

[Smith and Hancock, 2006] Smith, W. A. and Hancock, E. R. (2006). Recovering facial shape using a statistical model of surface normal direction. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 28(12):1914–1930.

[Stegmann et al., 2000] Stegmann, M. B., Fisker, R., and Ersbøll, B. K. (2000). *On properties of active shape models*. IMM, Department of Mathematical Modelling, Technical University of Denmark.

[Sukno et al., 2015] Sukno, F. M., Waddington, J. L., and Whelan, P. F. (2015). 3-D facial landmark localization with asymmetry patterns and shape regression from incomplete local features. *Transactions on Cybernetics*, 45(9):1717–1730.

[Sun et al., 2015] Sun, Y., Dong, J., Jian, M., and Qi, L. (2015). Fast 3d face reconstruction based on uncalibrated photometric stereo. *Multimedia Tools and Applications*, 74(11):3635–3650.

[Tewari et al., 2017] Tewari, A., Zollhöfer, M., Kim, H., Garrido, P., Bernard, F., Pérez, P., and Theobalt, C. (2017). Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. In *International Conference on Computer Vision (ICCV)*, volume 2, page 5. IEEE.

[Tran et al., 2017] Tran, A. T., Hassner, T., Masi, I., and Medioni, G. (2017). Regressing robust and discriminative 3d morphable models with a very deep neural network. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1493–1502. IEEE.

[Tzou et al., 2014] Tzou, C.-H. J., Artner, N. M., Pona, I., Hold, A., Placheta, E., Kropatsch, W. G., and Frey, M. (2014). Comparison of three-dimensional surface-imaging systems. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, 67(4):489–497.

[Tzou and Frey, 2011] Tzou, C.-H. J. and Frey, M. (2011). Evolution of 3d surface imaging systems in facial plastic surgery. *Facial Plastic Surgery Clinics*, 19(4):591–602.

[van Ginneken et al., 2003] van Ginneken, B., de Bruijne, M., Loog, M., and Viergever, M. A. (2003). Interactive shape models. In *Medical Imaging*, pages 1206–1216. International Society for Optics and Photonics.

[Vávra et al., 2017] Vávra, P., Roman, J., Zonča, P., Ihnát, P., Němec, M., Kumar, J., Habib, N., and El-Gendi, A. (2017). Recent development of augmented reality in surgery: a review. *Journal of healthcare engineering*, 2017.

[Veitch et al., 2012] Veitch, D., Burford, K., Dench, P., Dean, N., and Griffin, P. (2012). Measurement of breast volume using body scan

technology (computer-aided anthropometry). *Work*, 41(Supplement 1):4038–4045.

[Vlasic et al., 2005] Vlasic, D., Brand, M., Pfister, H., and Popović, J. (2005). Face transfer with multilinear models. *Transactions on graphics (TOG)*, 24(3):426–433.

[Woodham, 1980] Woodham, R. J. (1980). Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):191139.

[Xiong and De la Torre, 2013] Xiong, X. and De la Torre, F. (2013). Supervised descent method and its applications to face alignment. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 532–539. IEEE.

[Yan et al., 2016] Yan, X., Yang, J., Yumer, E., Guo, Y., and Lee, H. (2016). Perspective transformer nets: Learning single-view 3d object reconstruction without 3d supervision. In *Advances in Neural Information Processing Systems*, pages 1696–1704.

[Zeng et al., 2017] Zeng, D., Zhao, Q., Long, S., and Li, J. (2017). Examplar coherent 3d face reconstruction from forensic mugshot database. *Image and Vision Computing*, 58:193–203.

[Zhang and Besier, 2016] Zhang, J. and Besier, T. F. (2016). Accuracy of femur reconstruction from sparse geometric data using a statistical shape model. *Computer Methods in Biomechanics and Biomedical Engineering*, pages 1–11.

[Zhang et al., 2003] Zhang, L. et al. (2003). Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multiview stereo. In *International Conference on Computer Vision (ICCV)*, pages 618–625. IEEE.

[Zhang et al., 1999] Zhang, R., Tsai, P.-S., Cryer, J. E., and Shah, M. (1999). Shape-from-shading: a survey. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 21(8):690–706.

[Zhao et al., 2004] Zhao, M., Li, S. Z., Chen, C., and Bu, J. (2004). Shape evaluation for weighted active shape models. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pages 1074–1079. Citeseer.

[Zhao and Chellappa, 2000] Zhao, W. Y. and Chellappa, R. (2000). Illumination-insensitive face recognition using symmetric shape-from-shading. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 286–293. IEEE.

[Zhao and Chellappa, 2001] Zhao, W. Y. and Chellappa, R. (2001). Symmetric shape-from-shading using self-ratio image. *International Journal of Computer Vision*, 45(1):55–75.

[Zheng et al., 2006] Zheng, G., Rajamani, K. T., Zhang, X., Dong, X., Styner, M., and Nolte, L.-P. (2006). Kernel regularized bone surface reconstruction from partial data using statistical shape model. In *International Conference of the Engineering in Medicine and Biology Society (EMBC)*, pages 6579–6582. IEEE.

[Zhou et al., 2013] Zhou, F., Brandt, J., and Lin, Z. (2013). Exemplar-based graph matching for robust facial landmark localization. In *International Conference on Computer Vision (ICCV)*, pages 1025–1032. IEEE.

[Zhu et al., 2016] Zhu, X., Lei, Z., Liu, X., Shi, H., and Li, S. Z. (2016). Face alignment across large poses: A 3d solution. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 146–155. IEEE.

[Zhu et al., 2015a] Zhu, X., Lei, Z., Yan, J., Yi, D., and Li, S. Z. (2015a). High-fidelity pose and expression normalization for face recognition in the wild. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 787–796. IEEE.

[Zhu et al., 2015b] Zhu, X., Yan, J., Yi, D., Lei, Z., and Li, S. Z. (2015b). Discriminative 3D morphable model fitting. In *International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, volume 1, pages 1–8. IEEE.